Deep Neural Networks in Hearing Devices

Research in cognition, listening, and AI is reshaping the future of hearing aid processing

By LARS BRAMSLØW, PhD, and DOUGLAS L. BECK, AuD

For many hearing care professionals, the term "Deep Neural Networks" (DNNs) is a relatively new term. However, DNNs have already been successfully incorporated into many disciplines, are used by many professionals, and can be found in an astounding array of technologies. Applying DNNs to hearing aids represents an entirely new sound processing approach and technology. In our profession, with regard to hearing aids, DNNs offer the promise of higher sound quality, an improved signal-to-noise ratio (SNR), improved understanding of speech in noise, improved recall, and more.

Lars Bramsløw, PhD, is a Senior Scientist at the Eriksholm Research Centre (Oticon A/S) in Snekkersten, Denmark, who has worked and published extensively in the area of deep neural networks and hearing devices. **Douglas L. Beck, AuD,** is the Vice President of Academic Sciences at Oticon Inc, Somerset, NJ.

Although hearing aid amplification has certainly been successful with regard to making sounds audible and improving the quality of life for millions of people with hearing loss, the ability to clearly understand speech in noise (SIN) remains problematic for many. In fact, the inability to understand SIN is the primary complaint of people with sensorineural hearing loss (SNHL), and for many users of traditional hearing aids.

SIN complaints are generally founded within the very nature and definition of SNHL. As outer and inner hair cells are damaged or destroyed due to aging, trauma, noise, ototoxic chemicals and more, the quantity and quality of the neural code sent from the inner ear to the brain is diminished. Over time, as hearing loss increases, the neural code degrades even further. Thus, the brain must increase listening effort and expend more energy to make sense of, untangle, comprehend, or understand the degraded neural code. DNNs may facilitate an improved neural code—one that is easier to make sense of.

Traditional hearing devices have addressed the SIN challenge using advanced signal processing including directional microphones and noise reduction algorithms. For example:

- *Directional microphones* filter out or reduce steady-state noise and reverberation coming from behind the listener, while often assuming the signal of interest is in front of the listener. The drawback is the loss of situational awareness, distortion of interaural loudness and other localization cues, and a decreased perception of the entire auditory scene. Of note, if the disturbing noises are in front of the listener, the directional system cannot attenuate them.
- *Noise reduction algorithms* can suppress steady-state noises and provide increased comfort and reduced listening effort, but generally do not improve speech intelligibility in difficult listening situations, such as low signal-to-noise ratios (SNRs), high reverb environments, cocktail parties/restaurants, and
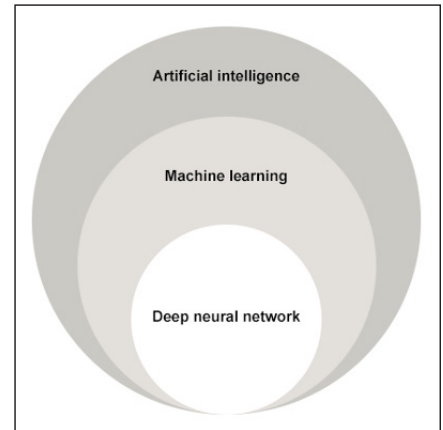

Figure 1. Deep Neural Networks (DNNs) are a subset of Machine Learning (ML), which is a subset of Artificial Intelligence (AI). DNNs rapidly analyze and interpret huge data sets. DNNs are developed to teach computers, processors, and other systems to respond (more-or-less) in a way similar to how a human might respond to vast quantities of incoming data, while incorporating constant checking, re-checking, correcting, and verifying.

other challenging acoustic environments.

Artificial Intelligence, Machine Learning, and Deep Neural Networks

The terms "artificial intelligence" (AI), "machine learning" (ML), and "deep neural networks" (DNN) can seem ambiguous and interchangeable. However, increasing levels of computational analysis, decision making, and sophistication occur as we traverse from the rather-all-inclusive term "AI" and progress to "ML," until we arrive at the most sophisticated and advanced protocol, DNN (**Figure 1**).

Moolayil[1] reports AI is the quality of intelligence introduced to a machine. To make machines "smarter" we facilitate machine-made choices. For example, washing machines use the right amount of water and the right temperature to soak, wash, and spin. Likewise, the thermostat in your refrigerator automatically adjusts the temperature within pre-selected parameters. The intelligence programmed into these relatively simple machines is referred to as Artificial Intelligence.

Machine Learning is more sophisticated than AI. Machine Learning "induces intelligence" into a machine without explicit programming.[1] Although the machine is not

encoded with a comprehensive list of permutations or rules, the machine learns—or is *trained*—from previous exposures. Copeland[2] reports machine learning uses algorithms to parse data, learn from it, and decide/predict something. For example, machine learning is used when a hearing aid switches from omni to directional settings when the hearing aid detects loud steady-state noise.

DNNs have been evolving for decades. DNNs are designed to make sense of huge data sets[2] and "teach" computers to do more-or-less what humans would do.[3] However, when DNNs arrive at a decision, they self-check. As the input data-set increases, the output is continually verified against the input. Through "successive approximation," accuracy increases as the input data set increases, all without specific instructional algorithms.

Similarly, for DNNs within a living human brain, there are no prescribed or pre-printed instructions. The human brain receives vast quantities of incoming data (ie, sensory experience), and over time, learning occurs. For example, most babies learn to talk and walk without specific direction. Generally, babies receive sensory input from their five senses for a year or so, and without specific instruction, their brain figures out and engages the protocols of walking and talking. Progress toward the goals (walking and talking) occurs and success is eventually achieved through successive approximation, corrections, edits, checks, and cross-checks. DNNs are also engaged as birds learn to fly and fish learn to swim.

Behind the Scenes

The idea of a DNN within a computer chip is admittedly difficult to grasp. However, computer-based, pragmatic, and integrated DNNs are all around us. They have been used, and are particularly well-suited to, make sense of "big data." DNNs are currently being applied in many ways behind the scenes to allow computers to perform tasks which were once exclusively human tasks, such as:

- Automatic speech recognition
- Facial recognition
- Language translation
- Image enhancement
- Automatic handwriting identification
- Streamed music auto-select functions
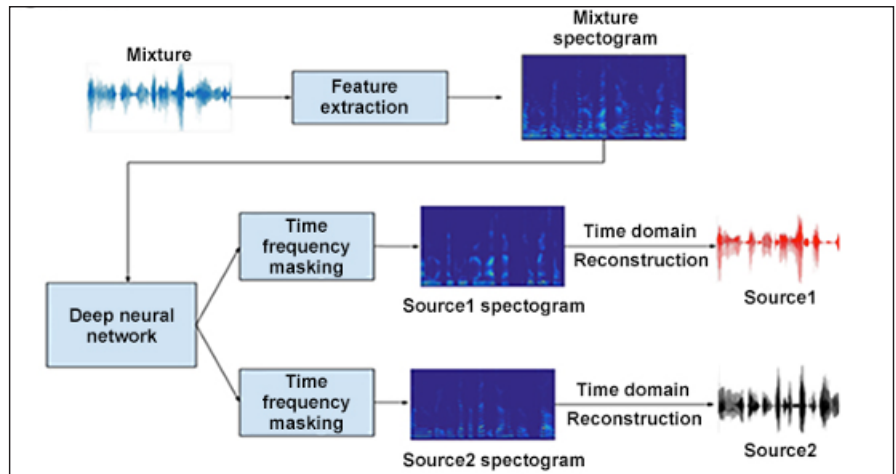- Drones delivering packages in crowded neighborhoods, and
- Self-driving cars.

Figure 2. Flowchart for a two-talker separation algorithm. In the case of two competing voices, two outputs can be directed to the two ears, while in a speech + noise scenario, "Source2" can be noise and thus discarded. Based on Naithani et al.[12]

At their core, DNNs are based on (relatively) simple calculations and represent "best-fit" data-driven, outcomes-based decisions.[4] DNNs are large computational networks of simple elements which are interconnected by many "weights"—similar to "synapses" in a brain. The weights are trained by presenting vast input examples and corresponding answers (outputs) to the network and slowly adjusting the weights, such that the network output achieves the best possible match to the input.

Initially, "training" the DNN is computationally intense. Training the DNN may depend on super-huge central computer network facilities and cloud services. However, once the DNN is trained, it requires much less power and can be installed into smart phones, hearing aids, and other devices. Indeed, the newest smartphones already include "neural co-processors."

DNN Speech Enhancement

One practical way of applying DNNs is via "time-frequency masks." In this application selected instances of the speech-in-noise signal are allowed, while the rest are suppressed. This is done by training the DNN to "ideal masks" that are derived by comparing "speech with noise" to "speech without noise" as the training target.[5,6]

The same approach can be used to identify and separate two voices, as demonstrated in **Figure 2**. The benefits of identifying and separating voices, with an emphasis on the primary speaker and attenuation of the secondary speaker, for hearing-impaired listeners has been documented across a range of conditions in laboratory and experimental settings.[7,8]

Regarding a future-based DNN speech enhancement feature, perhaps in years to come, one might envision a hearing aid trained to a specific voice or multiple voices. The selected voices would be encrypted and uploaded to train the DNN. The receiving computer would perform the personalized training and download the trained network weights ("model") to the hearing device. This would result in an improved speech-in-noise listening experience, as the hearing device is enabled to follow/track a specific voice or voices.

For example, in difficult listening situations such as a cocktail party or a restaurant, if the DNN was trained to attend to the conversation partner's voice, the SNR and overall quality of sound might be significantly enhanced. If there are multiple voices of interest present, the device would present them all to the listener, such that they can be more easily segregated and attended to via selective attention (paying attention to the person/voice the end-user chooses). These choices could be guided via a smartphone app, or

Regarding a future-base DNN speech enhancement feature, perhaps in years to come, one might envision a hearing aid trained to a specific voice or multiple voices.
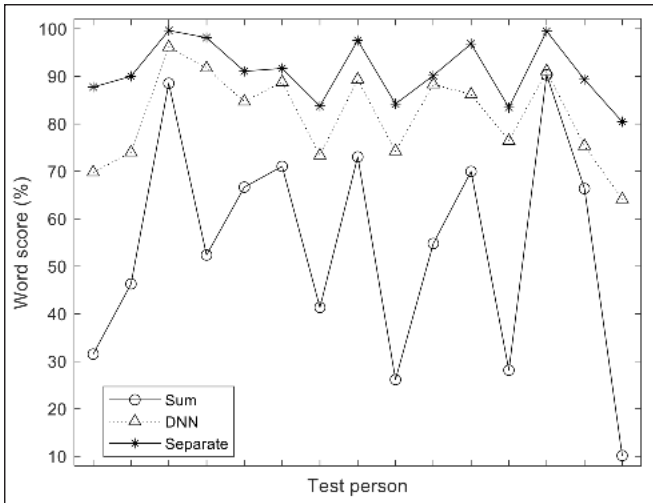
Figure 3. Word scores in a dual-talker competing voices test for the unprocessed sum, the DNN separated voices, and the (ideal) separated voices. Note the larger "DNN benefit" for the poorer "Sum" performers, and the smaller gap between the "DNN" and ideal separated voices (see Bramsløw et al[11] for details).

could be driven by the intent of the user[9] via a "brain-computer interface" guided by EEG signals from the ears,[10] or other beneficial indicators (eg, visual attention, etc).

The benefits of a two-competing-voices scenario was recently demonstrated by Bramsløw et al[11] using the DNN architecture proposed by Naithani et al.[13] On average a 13% increase in word recognition was obtained when attending to both voices, and more significantly, a 37% increase occurred when only one voice was selected. In general, those people who performed well initially received the smallest benefit from a voice separation algorithm, while people who perform poorly generally demonstrated more benefit (**Figure 3**). The same pattern occurs with degrees of hearing loss and cochlear implant users.[14] That is, more significant benefits are typically demonstrated for those with greater hearing difficulty.

Advanced DNN architectures are currently being researched whereby a longer history of speech sounds (ie, more input data) is used to augment the separation process.[12,14] Additionally, an algorithm focused on the most-relevant speech cues may prove useful.[15]

Greater Learning Capacity

DNN performance relies heavily on training with large data sets of representative stimuli, and DNN performance relies on learning, modifying, and predicting the preferred complex non-linear relations between input and output. Further, the hearing device could have detectors which enable the algorithm to engage only if improved SNR can be expected and delivered. For speech enhancement and for talker separation, millions of samples of speech and background noises are necessary, potentially facilitating an improvement in the SNR of approximately 2-10 dB, a very significant benefit for the hearing aid wearer.[16]

For an individualized and optimized DNN system, some degree of personal sound training will be beneficial. One option to obtain individualized and personalized training is to apply "transfer learning."[17] Transfer learning can be accomplished during product development and general updates sent to the hearing device via the internet, combined with personal and individual sound samples uploaded to the

cloud incorporated within the DNN, and then downloaded to the hearing device.

New Approaches to Signal Processing in Hearing Aids

The application of DNN in hearing devices represents an entirely new way of designing algorithms, and indeed, a whole new approach to sound processing. Traditional signal processing is "rule-based": detectors measure sound pressure levels, SNRs, and other acoustic properties of incoming sounds across multiple frequency bands, then these sounds trigger rules to calculate and update gain (and more) within the device. This can result in the suppression of sounds coming from a location in space not within the directional parameters selected, or detected as noise, or as too loud, or that have spectral content beyond the rules.

In contrast, DNN signal processing is "data-and-learning" based. DNNs create novel algorithms based on carefully selected training data. As a result of the DNN's data-and-learning based algorithm, the trained device behaves smoothly and intelligently across a vast range of sound environments in ways that traditional rule-based systems cannot.[18] Although a trained DNN is to some extent a "black box" which cannot be easily understood and interpreted, DNN hearing devices, such as the new Oticon More™, offer advantages in sound processing, sound quality, and speech-in-noise results which we believe professionals and end-users will find advantageous.

Hardware and Power: Moving Ahead to a Future with DNN

The original bench model DNN's were power-hungry. During general training, and to a lesser degree during application, these precursors to DNN-driven (on-the-chip) hearing aids were not practical to implement because of their power requirements.

However, commercially available applications are now benefiting from rapid technology development. Indeed, newer iPhones are already equipped with a neural co-processor which can efficiently run trained DNNs in real-time. For hearing aid applications, given advanced hearing aids with neural co-processors (on-the-chip or co-processing), only a fraction of the available hearing aid power is required for DNN processing. This development is commercially available in headsets and hearing aids using new chip designs.[19-21]

Cloud-based DNN systems are already in hearing devices. However, the new Oticon More includes what we believe to be the first embedded (on-the-chip) DNN, trained on 12 million real-life sound scenes, which is designed to provide superior noise reduction. DNNs offer contemporary, sophisticated and advanced sound processing. On-the-chip DNN technology in hearing aids offers a real-time sound processing approach, solution, and technology. DNNs offer the opportunity for higher sound quality, an improved SNR, improved understanding of speech in noise, better identification of speech and noise, improved recall, and more. All of which serves to deliver an improved hearing and listening experience to the end-user, from the hearing care professional.

CORRESPONDENCE can be addressed to Dr Beck at: dbec@oticon.com

REFERENCES can be accessed in the online version of this article at HearingReview.com.

References [online only]

1. Moolayil JJ. A layman's guide to deep neural networks. July 24, 2019. Available at: https://towardsdatascience.com/a-laymans-guide-to-deep-neural-networks-ddcea24847fb

2. Copeland M. What's the difference between artificial intelligence, machine learning and deep learning? July 29, 2016. https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

3. Mathworks. What is deep learning? 3 things you need to know. 2020. Available at: https://www.mathworks.com/discovery/deep-learning.html

4. Kelleher, JD. *Deep Learning*. Cambridge, Mass: MIT Press;2019.

5. Bentsen T, May T, Kressner AA, Dau T. The benefit of combining a deep neural network architecture with ideal ratio mask estimation in computational speech segregation to improve speech intelligibility. *PloS One.* 2018;2018 [May 15];13(5):e0196924. doi: 10.1371/journal.pone.0196924

6. Wang D, Chen J. Supervised speech separation based on deep learning: An overview. *IEEE/ACM Trans Audio Speech Lang Process.* 2018;99. doi:10.1109/TASLP.2018.2842159

7. Healy EW, Delfarah M, Vasko JL, Carter BL, Wang D. An algorithm to increase intelligibility for hearing-impaired listeners in the presence of a competing talker. *J Acoust Soc Am.* 2017;141:4230-4239. doi:10.1121/1.4984271

8. Healy EW, Delfarah M, Johnson EM, Wang D. A deep learning algorithm to increase intelligibility for hearing-impaired listeners in the presence of a competing talker and reverberation. *J Acoust Soc Am.* 2019;145:1378-1388. doi:10.1121/1.5093547

9. Beck DL, Pontoppidan NH. Predicting the intent of a hearing aid wearer. *Hearing Review.* 2019[Jun];26(6):20-21. Available at: www.hearingreview.com/hearing-products/hearing-aids/apps/predicting-intent-hearing-aid-wearer

10. Dau T, Maercher Roersted J, Fuglsang S, Hjortkjær J. Towards cognitive control of hearing instruments using EEG measures of selective attention. *J Acoust Soc Am.* 2018; 143(3), 1744-1744.

11. Bramsløw L, Naithani G, Hafez A, Barker T, Pontoppidan NH, Virtanen T. Improving competing voices segregation for hearing impaired listeners using a low-latency deep neural network algorithm. *J Acoust Soc Am.* 2018; 144(1):172-185. https://doi.org/10.1121/1.5045322

12. Naithani G, Barker T, Parascandolo G, Bramslow L, Pontoppidan NH, Virtanen T. Low latency sound source separation using convolutional recurrent neural networks. *2017 IEEE Work Appl Signal Process to Audio Acoust.* 2017:71-75.doi:10.1109/WASPAA.2017.8169997

13. Naithani G, Parascandolo G, Barker T, Pontoppidan NH, Bramsløw L, Virtanen T. Low-latency sound source separation using deep neural networks. *2016 IEEE Glob Conf Signal Inf Process.* 2016;272–276.

14. Goehring T, Keshavarzi M, Carlyon RP, Moore BCJ. Using recurrent neural networks to improve the perception of speech in non-stationary noise by people with cochlear implants. *J Acoust Soc Am.* 2019[Jul];146(1):705. doi: 10.1121/1.5119226.

15. Kolbæk M, Tan Z-H, Jensen SH., Jensen J. On loss functions for supervised monaural time-domain speech enhancement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing.* 2020;28:825-838.

16. Beck DL, Benitez L. A two-minute speech-in-noise test protocol and pilot data. *Audiology Today.* 2019; May/June. Available at: https://www.audiology.org/audiology-today-mayjune-2019/two-minute-speech-noise-test-protocol-and-pilot-data

17. Purwins H, Li B, Virtanen T, Schlüter J, Chang SY, Sainath T. Deep learning for audio signal processing. *IEEE Jour Select Topics Signal Processing.* 2019;13(2), 206-219.

18. Elberling C, Vejlby Hansen K. Hearing instruments-Interaction with user preference. Auditory Models and Non-linear Hearing Instruments In: *Proceedings of the 18th Danavox Symposium.* 1999;341-357. Available at: http://www.audiological-library.gnresound.dk/

19. Pontoppidan NH. Voice separation with tiny ML on the edge. TinyML summit 2020. Available at: https://www.tinyml.org/summit/abstracts/Pontoppidan_Niels_abstract.pdf

20. TinyML summit 2019. Available at: https://www.tinyml.org/summit/2019/index.html

21. TinyML summit 2020. Available at: https://www.tinyml.org/summit