



# Creating Clarity in Noisy Environments by Using Deep Learning in Hearing Aids

**Asger Heidemann Andersen, Ph.D.,<sup>1</sup> Sébastien Santurette, Ph.D.,<sup>1</sup>  
Michael Syskind Pedersen, Ph.D.,<sup>1</sup> Emina Alickovic, Ph.D.,<sup>2</sup>  
Lorenz Fiedler, Ph.D.,<sup>2</sup> Jesper Jensen, Ph.D.,<sup>1</sup> and Thomas Behrens, M.Sc.<sup>1</sup>**

## ABSTRACT

Hearing aids continue to acquire increasingly sophisticated sound-processing features beyond basic amplification. On the one hand, these have the potential to add user benefit and allow for personalization. On the other hand, if such features are to benefit according to their potential, they require clinicians to be acquainted with both the underlying technologies and the specific fitting handles made available by the individual hearing aid manufacturers. Ensuring benefit from hearing aids in typical daily listening environments requires that the hearing aids handle sounds that interfere with communication, generically referred to as “noise.” With this aim, considerable efforts from both academia and industry have led to increasingly advanced algorithms that handle noise, typically using the principles of directional processing and postfiltering. This article provides an overview of the techniques used for noise reduction in modern hearing aids. First, classical techniques are covered as they are used in modern hearing aids. The discussion then shifts to how deep learning, a subfield of artificial intelligence, provides a radically different way of solving the noise problem. Finally, the results of several experiments are used to showcase the benefits of recent algorithmic advances in terms of signal-to-noise ratio, speech intelligibility, selective attention, and listening effort.

**KEYWORDS:** noise reduction, beamforming, directionality, postfiltering

<sup>1</sup>Oticon A/S, Smørum, Denmark; <sup>2</sup>Eriksholm Research Centre, Oticon A/S, Snekkerten, Denmark.

Address for correspondence: (e-mail: [aand@demant.com](mailto:aand@demant.com)).

Hearing Aid Technology to Improve Speech Intelligibility in Noise; Guest Editor, Joshua M. Alexander, Ph.D.

Semin Hear 2021;42:260–281. © 2021. The Author(s).

This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-

License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Thieme Medical Publishers, Inc., 333 Seventh Avenue, 18th Floor, New York, NY 10001, USA,

DOI: <https://doi.org/10.1055/s-0041-1735134>.

ISSN 0734-0451.

Hearing aids are often misconceived as being simple amplifiers of sound. While this may have been true in the past, modern hearing aids use a vast array of technologies to help the user perceive their surroundings. One of these technologies, which particularly finds its usefulness in the most challenging and noisy environments, is the noise reduction system.

The primary “medicine” administered by a hearing aid is hearing loss compensation. This applies frequency-dependent gain, derived from the user’s pure-tone thresholds, and dynamic range compression to ensure that soft sounds are amplified enough to be audible without loud sounds being amplified so much as to cause discomfort or pain. However, despite such compensation, many users still report difficulty coping with noisy environments.<sup>1,2</sup> This suggests that the effects of hearing loss cannot simply be compensated away through the use of amplification.

While the origins of sensorineural hearing loss are complicated and incompletely understood, psychophysical experiments have revealed a range of deficits in the impaired hearing system that are not related to a simple loss of sensitivity. These include the following<sup>3</sup>:

- *Frequency spread of masking.* Noise present in one frequency region may spread over a broader range to disturb sounds in nearby frequency regions. This spread is more extensive for hearing-impaired listeners.
- *Temporal spread of masking.* Noise bursts may mask following sounds. The duration across which this effect is present tends to be longer for hearing-impaired listeners.
- *Reduced ability to use spatial cues.* This deficit reduces the ability to localize sound sources and the ability to improve speech understanding in noise via spatially selective attention.

The aforementioned deficits, which cannot be compensated by gain or compression, can make speech intelligibility in noisy environments worse. Therefore, hearing loss is often modeled as the sum of an attenuation component that can be compensated by amplification and additional distortion components that cannot.<sup>4,5</sup>

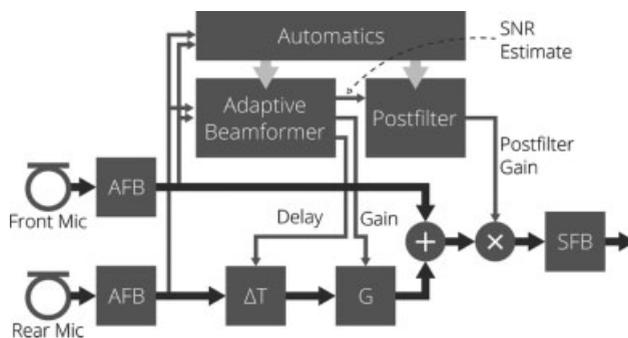
To reduce the impact of the deficits mentioned earlier and to make challenging listening environments more accessible to the user, modern hearing aids apply noise reduction algorithms. These tackle the difficulty of noisy environments directly by attempting to reduce distracting background noise without removing target sounds such as speech.

This article provides the reader with an understanding of the techniques used for reducing unwanted environmental noise in hearing aids. The focus will be on building intuition rather than on providing complete mathematical detail. Section 2 describes the typical structure of a noise reduction system as employed in a hearing aid. Such a system primarily comprises an adaptive beamformer, which removes noise by adapting the directional response of the hearing aid, coupled with a postfilter, which removes noise by applying time- and frequency-dependent attenuation to the signal. Section 3 describes how deep learning, a subdiscipline of artificial intelligence, is currently making completely new approaches for noise reduction available. After building basic intuition about the principle of deep learning, it is described how a neural network can be trained to replace the postfilter in a noise reduction system. This is shown to give rise to considerable improvements in noise reduction performance. Section 4 is a brief comment on the importance of using an automatic system to regulate the noise reduction system. Section 5 presents results from a selection of measurements and clinical studies that highlight the importance and continued improvement of noise reduction technology. Section 6 concludes upon the findings.

## 1: THE PRINCIPLES OF NOISE REDUCTION

This section provides an intuitive description of the core principles used for noise reduction in hearing aids. Fig. 1 shows the main components involved in such a noise reduction system. Two separate—but highly co-dependent—methods are used to reduce noise:

- **Beamforming** utilizes the fact that modern hearing aids most often have multiple microphones to amplify or suppress sounds



**Figure 1** An overview of the components used in the noise reduction system of a typical modern hearing aid. The signals from two microphones are converted to a time–frequency representation using separate analysis filterbanks (AFBs). An adaptive beamformer controls the directional response of the system by applying variable gains and time delays to one of the two signals before these are summed together. A postfilter computes a time- and frequency-dependent gain which is applied to the signal before a synthesis filterbank (SFB) converts the time–frequency representation of the signal back to an audio waveform.

depending on the direction from which they originate. This principle may also be referred to as directionality, directional processing, or spatial processing.

- **Postfiltering** aims at suppressing any leftover noise from the beamforming process. It does so by attenuating time–frequency regions that are dominated by leftover noise. Postfiltering is closely related to single-channel noise reduction.

Note that the term *noise reduction* is used here to refer to the joint use of these two principles, whereas some authors denote only postfiltering as *noise reduction* or *single-channel noise reduction*.

Here, the necessary concept of filterbanks is covered briefly (Section 2.1). Then beamforming (Section 2.2) and postfiltering (Section 2.3) are covered separately. Lastly, Section 2.4 comments on the strong integration between beamforming and postfiltering, both in theory and practice.

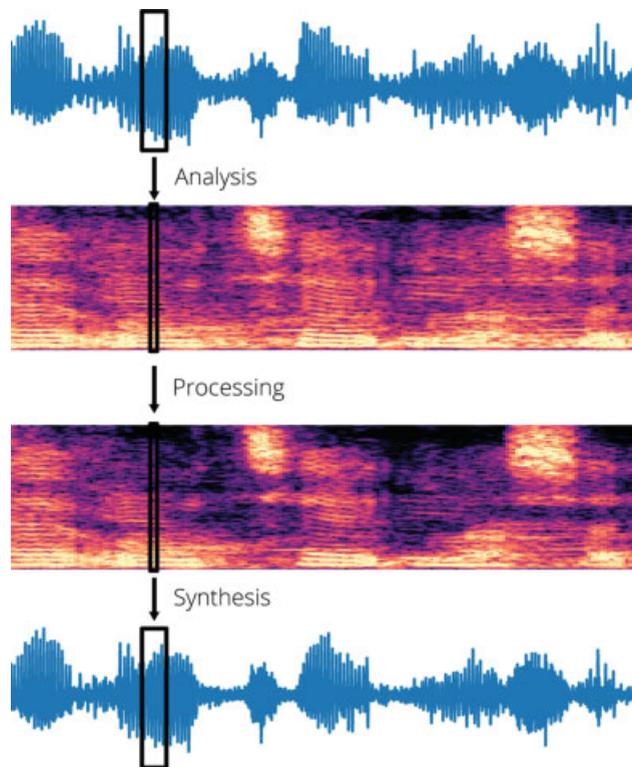
### 1.1: Analysis and Synthesis Filterbanks

The human auditory system has an amazing ability to discern different frequencies contained in audio signals.<sup>6</sup> Similarly, hearing aids can benefit from separately processing different frequency bands. The frequencies contained in an audio signal are, however, not readily visible from the raw audio waveform. This makes the raw audio waveform difficult to work with in

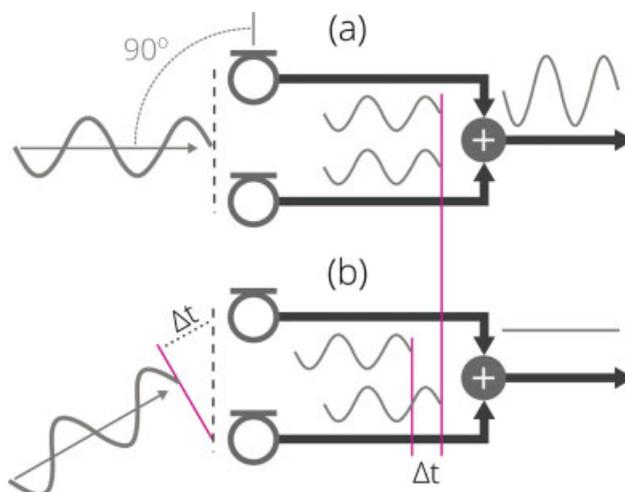
practice. Hearing aids, therefore, employ an analysis filterbank to split the input signal into short overlapping time segments and analyze the frequency content of these. This results in a signal representation that is closely related to a spectrogram. Most processing (e.g., beamforming and postfiltering) is conveniently performed on this signal representation. When the signal has been processed, a synthesis filterbank converts the signal back to an audio waveform by resynthesizing overlapping wave segments and combining them. The principle of analysis, processing, and synthesis is illustrated in Fig. 2.

### 1.2: Beamforming

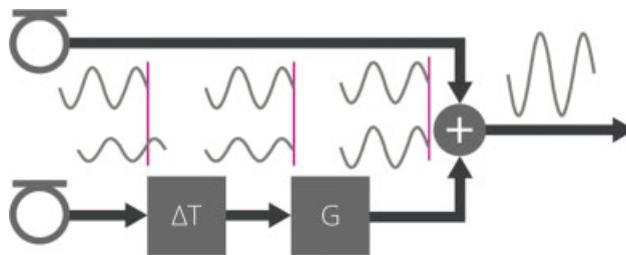
Modern hearing aids typically have two microphones mounted with a distance of approximately 6 to 12 mm, depending on the hearing aid style and brand. Depending on the direction of the impinging sound, it may arrive at one microphone slightly before the other. While this time difference is tiny (at most  $\sim 35$  microseconds), it holds valuable information about the direction of the sound. For instance, as Fig. 3 illustrates, if the two microphone outputs are simply summed together, the amplitude of the resulting signal depends greatly on the direction from which the sound arrived. This suggests that by simply summing the microphone signals, one can perform filtering in space: signals from certain directions can be suppressed completely, while signals from other directions can pass through unaltered.



**Figure 2** First, an analysis filterbank reveals the frequency structure inherent in an audio waveform of speech. Processing is performed in this representation, after which a synthesis filterbank is used to transform the result back to an audio waveform.



**Figure 3** The physical principle utilized in beamforming. (a) A single-tone signal impinging on a pair of microphones at an angle of 90 degrees relative to the axis of the microphones. The oscillations are picked up simultaneously by the microphones, resulting in signals that are in phase. When the two signals are summed, they add constructively to form a signal with twice the individual amplitude. (b) The signal impinges from a larger angle. Because of this, the sound arrives slightly earlier at the rear microphone compared with the front microphone. This causes the two signals to be out of phase. When summed, the signals cancel due to destructive interference.



**Figure 4** Showing how the principle illustrated in Fig. 3 can be controlled. The two microphones pick up signals that are not in phase and do not have the same amplitude. By applying a time delay and a gain to one of the signals, these differences are removed. The resulting signals sum constructively to a signal with twice the amplitude, even though the signals picked up by the microphones would not have.

A beamformer controls this phenomenon by applying additional gains and time shifts to one or both of the signals before summing them together. These parameters can be determined mathematically to ensure that sounds from specific directions are attenuated while sounds from other directions remain unaltered (see Fig. 4).

Beamforming allows an enormous degree of flexibility for continuously reconfiguring the directional properties of the hearing aid according to the current listening environment or the desired focus of the user. Hearing aids may offer a range of fixed directional patterns as well as adaptive directional patterns that change continuously to suit the environmental characteristics.

#### 1.2.1: FIXED BEAMFORMING

By determining appropriate fixed values for the delay and gain parameters applied in Fig. 4, it is possible to produce a range of static directional patterns, examples of which are shown in Fig. 5. The most straightforward of these is the omnidirectional response, which is produced by a single microphone, that is, by applying a gain of 0 ( $-\infty$  dB) to the other microphone. The omnidirectional pattern has the same sensitivity to all impinging sounds. It is typically preferred in environments where background noise is not an issue because it maintains the natural balance of the listening environment. The remaining patterns are left–right symmetric and have at most two *spatial nulls*, which are directions where sound is completely canceled. The dipole cancels sound from the sides while passing sound from the front and rear. The cardioid

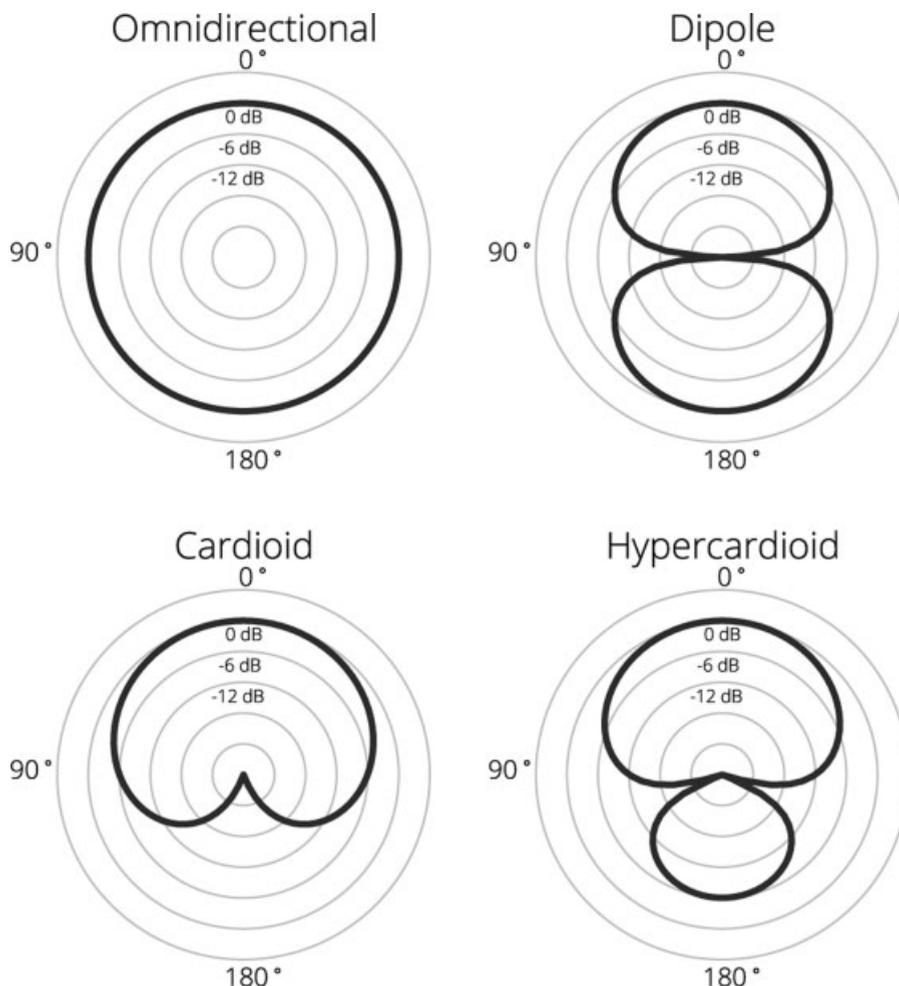
cancels sound from behind, making it particularly useful in listening environments where the target is located in the front and noise in the back. The hypercardioid has nulls placed at  $\pm 109$  degrees and provides the highest possible amount of noise reduction, assuming that the target is located in the front and the noise comes evenly from all directions (i.e., a *spherically diffuse noise field*). Please refer to Elko<sup>7</sup> for a thorough overview of the properties of various directional patterns.

The patterns shown in Fig. 5 assume free field acoustics and thus neglect the acoustic influence of the hearing aid shell and the user's head and body. The user's head has a considerable influence on the directional pattern that is actually realized, making it less symmetric by attenuating sounds coming from the opposite side of the head (see Fig. 6i in the article by Derleth et al in this issue for an example of this phenomenon).

#### 1.2.2: ADAPTIVE BEAMFORMING

Fixed beamformers force the user to either listen with the same directional pattern in all listening environments or make a conscious effort to change programs whenever a different directional response is desired. A less manual approach is to automatically adapt the beamformer parameters to minimize background noise across changing listening environments. Modern hearing aids tend to include at least some degree of adaptive beamforming, even in their default configurations.

A common approach for adaptive beamforming is the adaptive minimum variance distortionless response (MVDR) beamformer.<sup>8,9</sup>



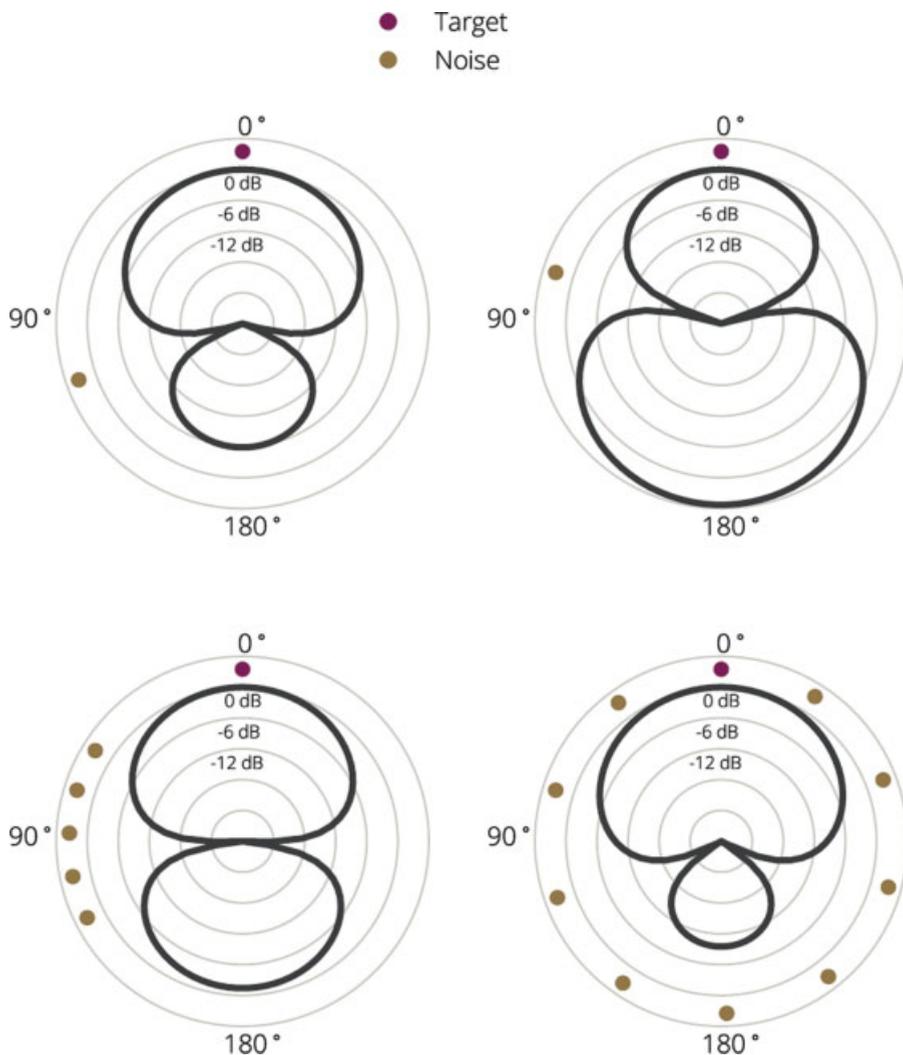
**Figure 5** Examples of directional responses that can be achieved using the described principles of beamforming. The plots show the attenuation of sounds reaching the hearing aid depending on the angle of arrival in the horizontal plane.

This collects statistics about the listening environment to derive beamformer parameters that (1) attenuate the total received sound as much as possible (i.e., achieve *minimum variance*), while (2) ensuring that sounds from the target direction are not attenuated or amplified (i.e., achieve a *distortionless response* toward the target). The target direction must be estimated separately or simply assumed to be directly in front of the user. Fig. 6 shows several examples of directional patterns arising from the use of an MVDR beamformer for different configurations of noise sources. The top left example shows how the MVDR beamformer can completely cancel a single noise source by placing a null in that direction. The bottom left example shows how

a group of noise sources can be attenuated by placing a null in the middle of them. The bottom right example shows the pattern that arises when several noise sources are distributed uniformly around the user.

The top right plot in Fig. 6 shows that, while the MVDR beamformer guarantees 0 dB gain in the target direction, it may actually amplify signals from other directions. Note, however, that this has no impact in this particular example since neither target nor noise is located in the directions with positive gain.

Since beamforming is applied to the frequency decomposition given by the analysis filterbank, different directional patterns can be applied for each frequency band. This allows



**Figure 6** Examples of directional responses achieved with an adaptive MVDR beamformer for different configurations of target and noise. In all four examples, the target is located in front of the user ( $0^\circ$ ), while one or more noise sources are located at directions indicated by the dots near the perimeter of the plots.

the adaptive beamformer to choose independent directional patterns that suppress the dominating noise sources in each frequency band.

MVDR beamforming is a very powerful technique to reduce background noise. However, for this same reason, it is often perceived as being too aggressive. Removing too much background noise can cause the user to feel detached from their surroundings. Therefore, such techniques require additional controls and limitations to be useful in practice. For instance, one might constrain the beamformer to select

only from “softer” patterns that do not have nulls, or avoid strict assumptions on where the target is located.

### 1.3: Postfiltering

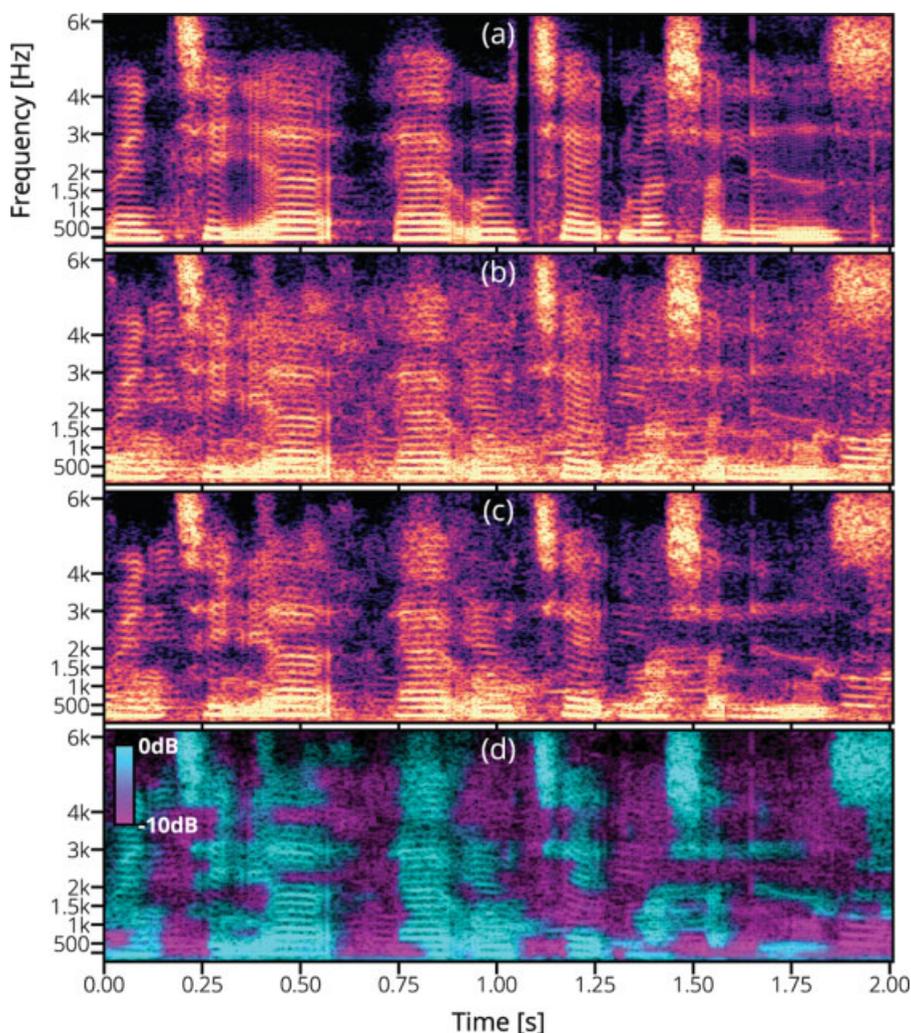
Beamforming is a very powerful tool for removing background noise whenever speech and noise arrive from different directions. It is, however, unable to remove noise from the target direction. This problem can instead be approached using methods from single-microphone noise reduction. When such processing

is applied after beamforming, it is often referred to as postfiltering. Such methods attempt to attenuate time–frequency regions in the signal (as seen in a spectrogram) dominated by noise. They do so by applying a postfilter gain of less than 0 dB to noisy regions. The most well-known of these methods, the Wiener filter,<sup>10</sup> uses a time-varying estimate of the signal-to-noise ratio (SNR) in each frequency band to suppress noise at times and frequencies where this can be done with little effect on the target signal. Mathematically, the method aims to make the filtered time-domain signal as similar to the target signal as possible (in a *mean squared error* sense). Other methods typically operate

according to a similar principle, but they aim to solve slightly different mathematical problems or rely on different speech and noise models.<sup>11,12</sup>

The processing of a postfilter is most easily visualized by considering a spectrogram of noisy speech, such as Fig. 7b. A good postfilter would suppress all noise-dominated time–frequency regions, leaving the speech unharmed. If done well, the result should be similar to the clean speech shown in Fig. 7a.

If the underlying target signal is known (as it is when imagining what a good postfilter should do to Fig. 7b while observing Fig. 7a), such processing can be almost infinitely



**Figure 7** (a) A spectrogram of a speech utterance. (b) The same utterance mixed with 24-talker babble at +3 dB SNR. (c) The noisy utterance after postfiltering. (d) Gray scale version of **b**, colored according to the gain applied by the postfilter.

effective. For instance, Kjems et al<sup>13</sup> showed that noisy speech at  $-60$  dB SNR can be rendered completely intelligible by such processing.

In real-world scenarios, as faced by hearing aid users, the target signal is obviously not known (one might even ask, “*why attempt to remove the noise if the underlying target signal is already known?*”). Postfiltering algorithms must instead rely on their own statistical estimates of the target and noise properties to determine which parts of the signal to attenuate. Fig. 7c shows the result of such processing, as applied by a typical hearing aid. In comparison with Fig. 7b, significant amounts of noise are clearly removed. On the other hand, some noise remains, and spectral and temporal details are smeared when comparing the postfiltered signal to the original target signal (Fig. 7a). Fig. 7d shows a spectrogram of the noisy signal, colorized according to the attenuation applied by the postfilter. This clearly reveals that the postfilter correctly applies attenuation (as shown in purple) in many regions with little or no speech while not attenuating (as shown in cyan) regions with mostly speech.

#### 1.4: Integrated Beamforming and Postfiltering

The previous sections have treated beamforming and postfiltering as two separate techniques, postfiltering being essentially just single-channel noise reduction applied to the beamformer output. There are, however, important links between the two systems. As noted, the Wiener filter attempts to filter a single noisy signal to make it resemble the target signal as closely as possible. The same mathematical problem can be formulated when multiple microphones are available. The solution to this problem is known as a multichannel Wiener filter.<sup>14</sup> It can be shown to be mathematically identical to an MVDR beamformer coupled with a single-channel Wiener filter.<sup>15</sup> Thus, the combined use of beamformers and postfilters for noise reduction is a theoretically optimal strategy—it arises as a mathematical consequence when solving the noise reduction problem.

A related fact makes the combined use of beamformers and postfilters even more interesting. As stated, the postfilter requires statistical

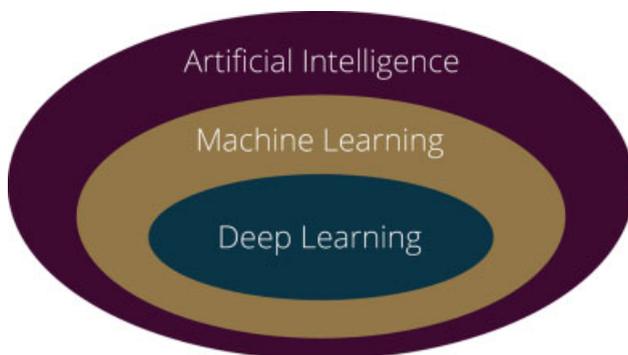
estimates about the target and noise, which are used to decide when and where to attenuate. For a Wiener filter, this involves estimating the short-time SNR in each frequency band. The beamformer is uniquely suited to help with the accurate estimation of SNR.<sup>16,17</sup> While a single directional pattern must be chosen for processing the signal to be presented to the user, nothing prevents the hearing aid from simultaneously using multiple other directional patterns for the explicit purpose of accurately estimating SNR<sup>16–18</sup> (see the article by Jespersen et al in this issue for a similar approach that uses dual microphones to estimate noise levels). This represents a significant difference between single-channel noise reduction and postfiltering.

Researchers have often found that single-channel noise reduction has no impact on, or may even deteriorate, speech intelligibility.<sup>19–21</sup> This turns single-channel noise reduction into a tradeoff between speech intelligibility and listening comfort. This result is often mistakenly extended to postfiltering. However, because noise reduction relies on accurate estimates of SNR and because beamformers can help provide these, postfiltering has a significant advantage compared with single-channel noise reduction. In practice, postfiltering can therefore increase speech intelligibility, even in normal-hearing listeners.<sup>22</sup>

## 2: NOISE REDUCTION USING MACHINE LEARNING

Throughout the last decade, artificial intelligence has transformed many technologies beyond recognition, including hearing aids (see the articles by Fabry and Bhowmik and by Balling et al in this issue for additional applications of artificial intelligence to hearing aids). These breakthroughs have mostly come from a subfield of machine learning called “deep learning” (see Fig. 8), which covers the training and use of neural networks for solving tasks.<sup>23</sup> Neural networks with multiple layers are sometimes referred to as deep neural networks (DNNs). Like many other technologies, deep learning has already had an enormous impact on noise reduction technology.

The previous section covered noise reduction without reference to techniques that



**Figure 8** Deep learning refers to the training and use of neural networks to solve tasks. It is a subfield of machine learning which itself is a field of artificial intelligence.

employ machine learning or deep learning. The discussed classical methods are characterized by using statistical models and methods to tell the target signal and background noise apart. However, there is a limit to the accuracy with which such models can reflect the diversity of real-world listening environments. This is because the models need to be fairly simple to allow for carrying out the mathematical derivations that lead to noise reduction algorithms. For instance, it is common to assume that speech is not correlated across frequency, that is, that there is no correspondence between what happens at one frequency and what happens at another frequency at the same moment. However, speech signals contain an intricate phonetic structure that is indeed highly correlated across frequency. By assuming independence of frequency channels, noise reduction algorithms miss the opportunity of benefiting from the structure of speech.

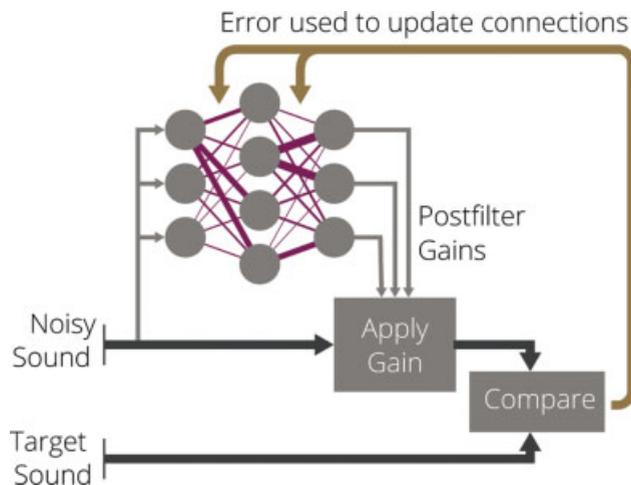
Machine learning (including deep learning) approaches the same problem in an entirely different manner. Instead of directly designing a specific algorithm to carry out a task (e.g., reducing noise), machine learning applies flexible, generic algorithms that can be trained to solve a task by analyzing examples of how the task should be solved. The applied algorithm is completely free to model whatever structures can be found in the examples, and there is no requirement for the solution to be mathematically simple or easy to explain. See Bishop<sup>24</sup> for a thorough overview of machine learning.

## 2.1: Training a Neural Network for Postfiltering

This section explains the basic principles involved in training a neural network to reduce noise. The training is executed on a database of examples of corresponding clean and noisy speech signals, such as the pair that comprise Fig. 7a and Fig. 7b. Pairs like these are referred to as training examples. The aim is to train a neural network to compute postfilter gains that make the noisy signals similar to the clean ones. The architecture used for doing so is shown in Fig. 9.

The neural network itself is composed of layers of neurons. The neurons in a layer are connected to the neurons in the previous layer by connections of varying strength.

An input to the neural network is a sequence of numbers: one number per neuron in the input layer. The input is transmitted and processed through the layers of neurons via the connections that link the layers. Finally, the last layer of the neural network returns an output, given as a sequence of numbers: one for each neuron in the output layer. Therefore, the neural network is simply a machine that takes an input and produces a corresponding output. How the output depends on the input is governed by a large number of parameters, given by the strengths of the connections between the layers. The number of parameters (connections) can range from thousands to billions depending on the design of the neural network (the famous GPT-3 language model trained by researchers at OpenAI has 175 billion parameters<sup>25</sup>). Training a neural network corresponds to adjusting the parameters in a way that makes the neural network solve a task.



**Figure 9** Showing how a neural network is trained to perform postfiltering. The neural network is used to compute postfilter gains for examples of noisy audio from the training database. These postfilter gains are applied to the noisy signals, and the result is compared with the underlying clean target signal using a loss function. Through the mathematical techniques of backpropagation and gradient descent, the neural network connections are updated to make the loss progressively smaller so that the postfiltered noisy signal is more similar to the underlying clean target.

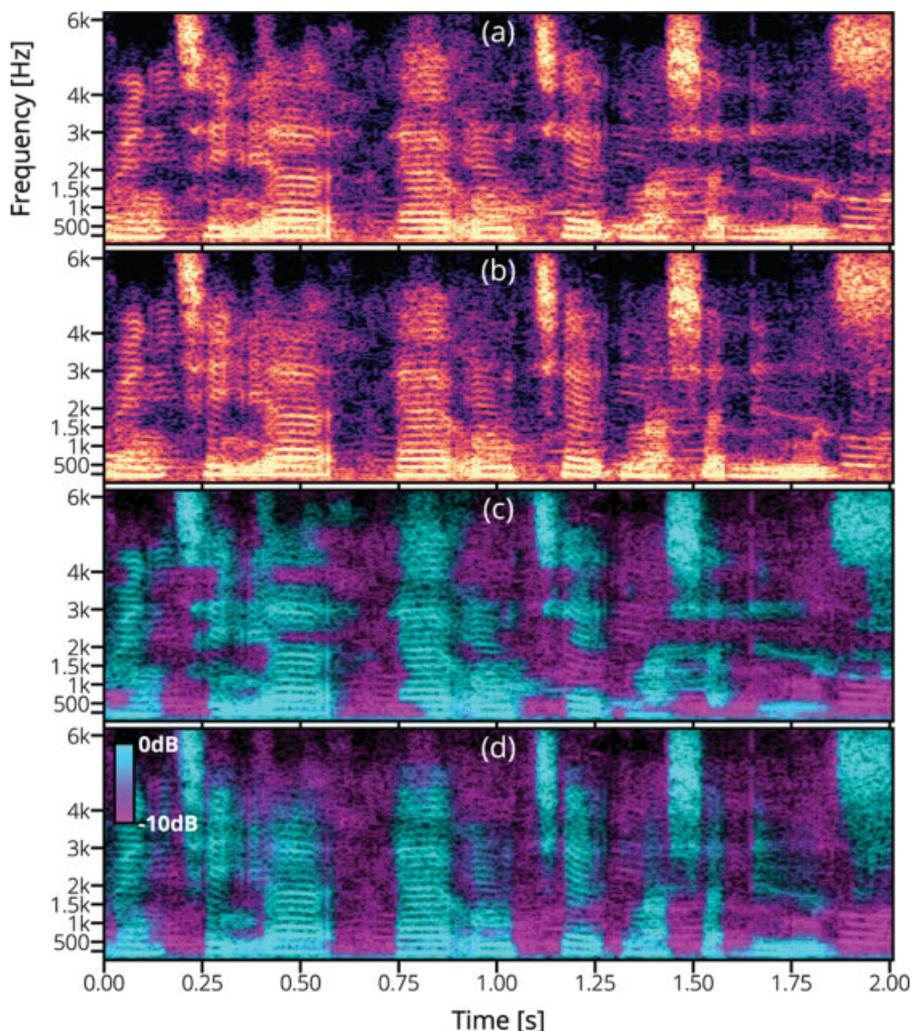
To use a neural network for postfiltering, an input that is somehow derived from the noisy signal is provided. This could correspond to simply the output of the beamformer or something more refined like estimates of SNR. The neural network outputs are the postfilter gains that are applied to the noisy signal (one gain value per frequency band).

Before training, the connections of a neural network are typically initialized to random values. Thus, to begin with, when a noisy signal is presented to the system, the neural network behaves mostly arbitrarily. The resulting, poorly postfiltered signal is compared with the target signal using a numerical rating known as a *loss function*. A loss function is a numerical metric that quantifies the difference between the two signals. For the untrained neural network, the loss function will likely report that there is a poor similarity between the postfiltered noisy signal and the target signal. The aim is to adjust the neural network connections through training to improve this similarity or, more specifically, decrease the loss.

Using a technique known as *backpropagation*, one can compute backward from the loss value to determine how a small change in any parameter would affect the loss. Using this knowledge, one can devise a small update to

all the neural network parameters, which will tend to slightly decrease the loss. When repeated over and over for different training examples, this process is known as *stochastic gradient descent*. If done carefully, this gradually causes the neural network to start behaving like a postfilter. Interestingly, this is achieved solely by showing the neural network examples of what a good postfilter should do (i.e., make the noisy signal less noisy), but without ever specifying how to do so.

Fig. 10a, b shows the output when the noisy signal from Fig. 7b is processed with a conventional postfilter and a postfilter based on a neural network, respectively. Processing with a neural network (Fig. 10b) results in a notably sharper and more speech-like result. This difference becomes even more apparent when comparing the applied postfilter gains, as shown in Fig. 10c, d. The conventional postfilter largely succeeds in identifying the speech regions, but otherwise appears somewhat uncoordinated. In contrast, the neural network postfilter displays a sharp and coordinated behavior across both time and frequency, correctly identifying most of the underlying speech and letting it through. These differences are not merely visual—the neural network postfilter



**Figure 10** Comparison of conventional postfiltering and DNN-based postfiltering. (a) A noisy speech utterance processed by a conventional postfilter (same as Fig. 7c). (b) The same noisy utterance processed by a DNN-based postfilter. (c) A gray scale spectrogram of the noisy utterance colorized according to the gain applied by the conventional postfilter (same as Fig. 7d). (d) Same as c, but for the DNN-based postfilter.

improves the speech intelligibility index (SII) weighted SNR by almost 2 dB over the conventional postfilter in the example shown.

While the above serves mainly as an illustration of the advantages associated with the use of neural networks for noise reduction, many academic studies have found comparable benefits on technical measures.<sup>26</sup> Behavioral studies have also reported intelligibility improvements in hearing-impaired listeners<sup>27–29</sup> and even normal-hearing listeners.<sup>30</sup> Similarly, it has been reported that normal-hearing listeners prefer neural network-based noise reduction

to conventional noise reduction.<sup>31</sup> There are, however, many intricacies involved in the training and evaluation of systems based on machine learning that can make it difficult to assess the real-world implications of such results. After carefully training and testing a state-of-the-art system based on neural networks to ensure that it was not evaluated on data that it had seen during training, Kolbæk et al<sup>26</sup> found that it could not reliably improve speech intelligibility for normal-hearing listeners. This result, however, was obtained for a single-channel noise reduction system, which generally does not

benefit from the improved SNR estimates that a directional system can produce.

## 2.2: Collection of Environmental Recordings

An essential resource for training neural networks is the database of training examples. Academic studies, which are most often focused on single-channel noise reduction, typically generate examples by mixing recordings from publicly available databases of speech and noise recordings. This allows large training databases to be produced while retaining complete control over factors such as noise type and SNR. On the contrary, such artificially produced sound examples are typically neither ecologically plausible nor representative of everyday environments for a hearing aid user. Furthermore, when training noise reduction systems for hearing aids, one relies on input signals as recorded from the hearing aid's microphones, including the acoustics of the hearing aid shell and the user's head. When training neural networks for noise reduction at Oticon, the authors have found that a good—albeit time-consuming—solution to the discussed issues is to use a database of ecologically valid spherical microphone array recordings. A substantial collection of such recordings has therefore been made. These consist of real conversations in different noisy listening environments commonly experienced by hearing aid users. The recordings were made at various physical locations, such as restaurants, cafés, offices, cars, and busy streets. The complete workflow from recording to training is illustrated in Fig. 11.

The sound environments were captured with a spherical microphone array containing 32 microphone capsules (Fig. 11a). This recording technique allows the sound environments to be reproduced in a sound studio with many loudspeakers. The sound-rendering procedure is described by Minnaar et al.<sup>32</sup> The technique relies on a calibration step where the microphone array is placed at the center of the loudspeaker array so that the transfer functions from all loudspeakers to all microphones on the sphere can be measured (Fig. 11b). Using an inverse filtering method,<sup>33</sup> each loudspeaker signal is computed as the sum of the microphone recordings that have been filtered to render the sound at the center of the loudspeaker

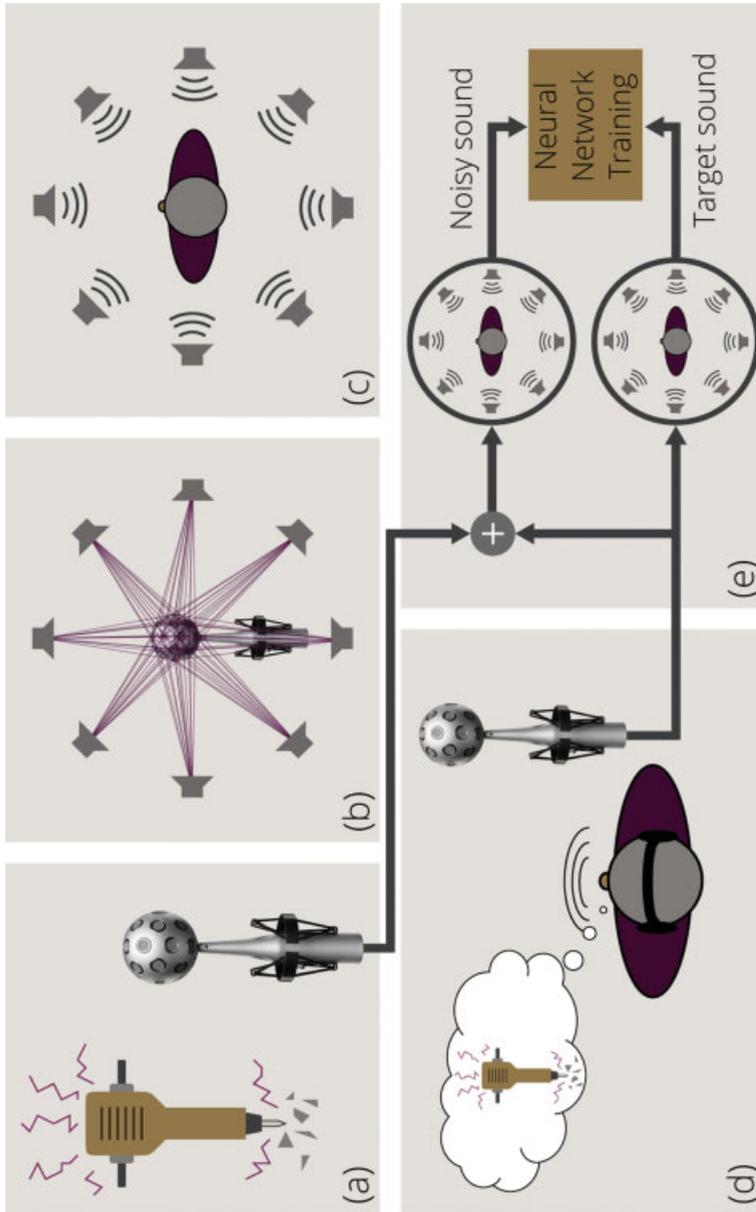
array as close as possible to the original sound recorded by the microphone array. With more loudspeakers, a better rendering of the original listening environment can be obtained.

With this approach, an acoustic scene of the original listening environment can be accurately reproduced near the center of the loudspeaker array (Fig. 11c). Before the acoustic scenes can be used as training material for neural networks, it is necessary to reproduce them as if they were recorded by a hearing aid mounted on a person's ear. A simple option could be to record from the microphones of a hearing aid mounted on a person or a manikin at the center of the loudspeaker array. However, to avoid the inconvenience of doing so for a large number of recordings, one can instead measure impulse responses from the studio loudspeakers to the hearing aid microphones. These can then be used to quickly accomplish the same result for any number of recordings, hearing aid styles, or people.

When using acoustic scenes as training material for a neural network, it is necessary to have separate recordings of the target speech signal and the background noise. It is well-known that humans tend to raise their vocal effort when speaking in a noisy background.<sup>34</sup> Therefore, an acoustic scene consisting of background noise mixed with a target talker who was recorded in the absence of noise will be perceived as unnatural because the vocal effort does not correspond to the noisy background. To improve the ecological validity of the acoustic scene, the original recording of the listening environment (Fig. 11a) is converted into a binaural audio signal. In the absence of noise, the target signal is recorded while the noise is presented to the talker(s) via open headphones (Fig. 11d). In this way, target speech and noise for a given acoustic scene are captured separately. Finally, the recorded speech and noise signals are mixed to generate an ecologically valid acoustic scene.

## 3: PERSONALIZATION AND AUTOMATICS

The noise reduction systems described in Sections 2 and 3 are highly effective at removing noise. However, at the same time, they



**Figure 11** The workflow involved in using spherical microphone array recordings for training neural networks. (a) Noisy listening environments are recorded with a spherical microphone array. (b) The microphone array is placed in the center of a loudspeaker array. The transfer functions from all loudspeakers to all microphones are measured. (c) Using techniques from Minnaar et al.<sup>32</sup> the transfer functions are inverted to reproduce the recorded listening environment at the center of the array. (d) Target audio is recorded by having one or more participants listen to noise recordings via open headphones while conversing in a quiet environment. (e) The acoustic scene is obtained by summing the noise and target recordings. Target and noisy sound signals are rendered to hearing aid microphones and used for neural network training.

introduce various forms of unwanted distortion. Furthermore, there is generally a large variation among hearing aid users regarding the preferred amount of noise reduction.<sup>19</sup> Such factors have led researchers to introduce heuristic limits that control the influence of the noise reduction system.<sup>35</sup> This makes it possible to mostly eliminate unwanted distortion and to adjust the amount of noise reduction to meet the user's preference.

The preferred amount of noise reduction varies across users, but it also varies across time. In a very noisy environment like a busy restaurant, most users may be willing to tolerate some distortion as long as the noise reduction provides the needed relief from the background noise. On the other hand, in a quiet environment, noise reduction might not be necessary or desired. Modern hearing aids have an *automatic* system that continuously adapts the noise reduction system to suit the listening environment. Automatic adjustment of the hearing aid is based on the results of an environmental classifier and the user's preferences for noise reduction as selected during the fitting process (see the article by Hayes in this issue for more details on environmental classifiers). The automatic system primarily acts by controlling the amount of directionality and postfiltering applied (as shown in Fig. 1), but it may influence other systems in the hearing aid too.

When surveying the academic literature on noise reduction, it becomes clear that the topic of automatic systems is an underappreciated part of hearing aid design. This is perhaps because it is a relatively softer discipline than the mathematically exact one of designing the underlying noise reduction system. However, the automatic system serves a critical function by ensuring that the individual user is exposed to the correct amount of noise reduction in any given listening environment. For the same reason, the clinician responsible for the fitting must be well-acquainted with the features of the noise reduction and automatic systems in the selected hearing aid.

#### 4: TECHNICAL AND CLINICAL BENEFITS OF NOISE REDUCTION

This section reports the results of technical and clinical investigations of the effects of different

noise reduction systems based on the approaches described in the previous sections, using two commercially available premium hearing aids (referred to as HA1 and HA2 in the following). HA1 employs a 16-channel noise reduction system with a fast-acting combination of an MVDR beamformer<sup>16</sup> and a single-channel Wiener postfilter.<sup>17</sup> HA2 employs a fast-acting 24-channel noise reduction system with a higher-resolution MVDR beamformer combined with the processing of a DNN-based postfilter that was trained to enhance the contrast between speech and noise using across-channel information.<sup>36</sup>

#### 4.1: Signal-to-Noise Ratio Benefit

To compare the SNR benefits of the two hearing aids, output SNR measurements were performed using the Hagerman and Olofsson phase-inversion technique<sup>37</sup> for HA1 and HA2. For each, a pair of hearing aids were fitted to the ears of a head-and-torso simulator (HATS) using closed-ear tips. The HATS was placed in the center of a circular loudspeaker setup containing 12 loudspeakers equally spaced by 30 degrees in the horizontal plane. Continuous speech was presented from the front loudspeaker at 0-degree azimuth, while cafeteria noise with an overall level of 65 dB SPL was presented from all loudspeakers simultaneously, such that noise came from all directions, including that of the speech, a situation that is especially challenging for traditional noise reduction systems. The measurements were obtained for speech levels of 60 dB SPL (corresponding to  $-5$  dB unaided

**Table 1** SII-weighted output SNR improvement in dB, relative to the unaided output SNR, for HA1 and HA2 at two different input SNRs when noise reduction is deactivated ("off"), the postfilter only is activated ("PF only"), and both beamformer and postfilter are activated ("BF + PF")

	-5 dB input SNR		0 dB input SNR	
	HA1	HA2	HA1	HA2
Off	-0.75	-0.16	-1.18	-0.39
PF only	0.11	1.81	-0.08	2.16
BF + PF	4.04	4.54	3.82	4.65

SNR) and 65 dB SPL (0 dB unaided SNR). The hearing aid output signals were recorded via the HATS internal microphones with the phase of the noise either unchanged or inverted so that the speech and noise levels at the output of the hearing aids could be estimated.<sup>37</sup>

The SII-weighted output SNR improvements using this method are listed in Table 1 for measurements in which (1) the noise reduction systems in HA1 and HA2 were disabled (“Off”); (2) only the postfilter algorithm was activated (“PF only”); and (3) both the beamformer and the postfilter were activated (“BF + PF”). Note that all output SNRs reported in Table 1 for these three conditions are relative to the unaided output SNR (similar to the input SNR), such that positive values reflect an SNR improvement and negative values reflect a worsening of the SNR.

In the “PF only” condition, the DNN-based HA2 produces SNR improvements that far exceed the Wiener-filter-based HA1 (a 1-dB increase in SNR can lead to an increase of 10 percentage points in speech intelligibility when performance is at the steepest portion of the performance-intensity curve). This SNR benefit will be partly or fully present in environments where the automatics system does not fully use beamforming. The results of “BF + PF” show that the full activation of beamforming provides an even larger SNR benefit that can exceed 4 dB. At the same time, the effects of beamforming somewhat reduce the postfilter-related differences between HA1 and HA2 in the “PF only” condition. While beamforming is highly effective, it should be noted that aggressive beamforming can lead to side effects such as feeling detached from one’s surroundings (see the articles by Jespersen et al; Derleth et al; and Branda and Wurzbacher in this issue for additional discussion about this problem). Therefore, users are rarely exposed to the full potential of beamforming.

#### 4.2: Speech Intelligibility Benefit

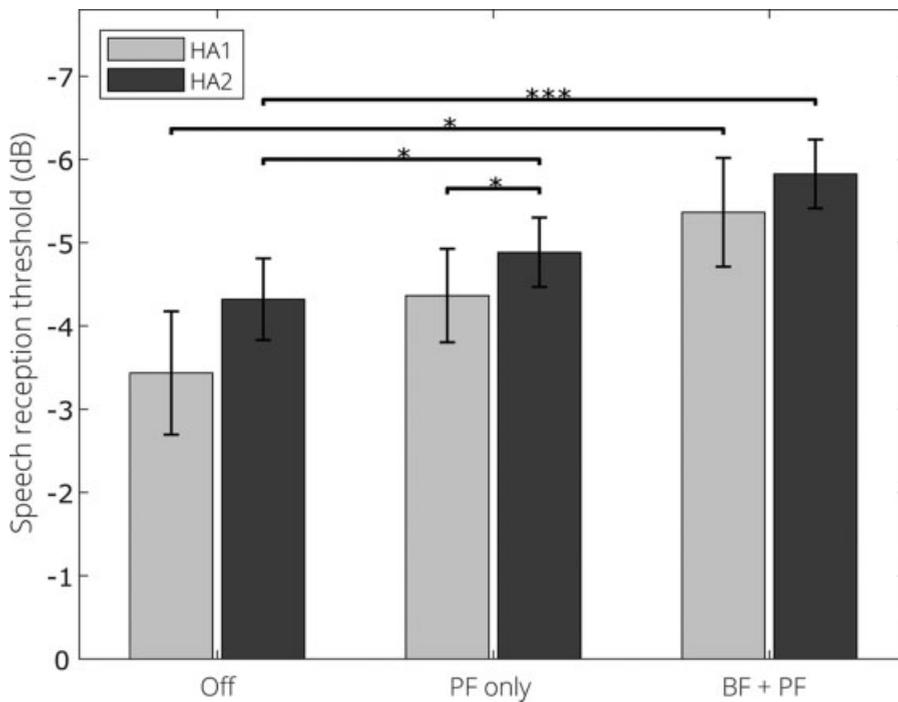
While technical benefits, like those described in the previous section, can be measured, there is no guarantee that these will translate into improvements in speech intelligibility. To test whether the documented SNR improvements provided by the DNN-based HA2 translate to

improved speech intelligibility in noise, 20 experienced adult hearing-aid users completed a matrix sentence test. Participants had mild-to-moderate symmetrical sensorineural hearing loss and qualified to be fit with receiver-in-the-ear (RITE) hearing aids. They also provided informed consent and received financial compensation for their participation. The current study was approved by the ethics committee of the University of Oldenburg.

All participants performed the Oldenburg sentence test<sup>38</sup> while wearing either HA1 or HA2 fitted with closed-ear tips and amplification based on the voice-aligned compression (VAC +) rationale, a quasi-linear fitting rationale with low compression knee-points based on the loudness data from.<sup>39</sup> The same test setup and stimuli as in the previous technical measurements were used, with an overall noise level of 68 dB SPL and an adaptive speech level. After performing two training lists, each participant’s speech reception threshold (SRT) for a 50%-correct intelligibility level was measured for each hearing aid in the “Off,” “PF only,” and “BF + PF” conditions (see Section 5.1). This yielded a total of six test conditions that were measured in random order.

Fig. 12 shows the mean SRTs obtained for each condition. On average, activating the different components of the noise reduction systems led to increased speech intelligibility (i.e., lower SRTs). Furthermore, HA2 (dark gray bars) led to higher intelligibility than HA1 (light gray bars). A repeated-measures analysis of variance (ANOVA) revealed significant main effects of hearing aid ( $[F_{1,19} = 5.1, p < 0.035]$ ) and noise reduction [ $F_{2,38} = 17.6, p < 0.001$ ]. Post hoc multiple comparisons using Tukey’s honest significant difference criterion showed that SRTs in the “BF + PF” and “Off” conditions differed significantly for both hearing aids (HA1:  $p = 0.022$ , HA2,  $p < 0.001$ ). The “PF only” versus “Off” comparison was significant only for HA2 ( $p = 0.036$ ). The only condition in which SRTs differed significantly between HA1 and HA2 was the PF only condition ( $p = 0.046$ ).

These results confirm that the investigated noise reduction systems’ SNR benefits translate into real speech intelligibility improvements in a complex listening environment. Note,



**Figure 12** Mean SRTs for 50% correct speech intelligibility obtained in the Oldenburg sentence test ( $N = 20$ ). Error bars indicate the standard error of the mean. Note that the  $y$ -axis is reversed, such that higher bars indicate higher speech intelligibility. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

especially, that the DNN-based HA2 in the “PF only” condition produces a statistically significant improvement in intelligibility compared with “Off.” This runs counter to the conventional expectation that only beamforming can improve intelligibility and clearly showcases the differences between postfiltering and single-channel noise reduction.

### 4.3: Effects on Cortical Representations and Listening Effort

Noise reduction systems in hearing aids have been shown to reduce listening effort during speech recognition tasks in noise (e.g., as shown by Ohlenforst et al<sup>40,41</sup>) and to enhance the cortical representation of speech in the auditory cortex in noisy multitalker environments.<sup>42,43</sup> The protocols from previous electroencephalography (EEG) and pupillometry studies<sup>42,44,45</sup> were adapted to compare how the noise reduction systems from HA1 and HA2 affect these two outcomes. Since the same protocols were strictly followed, only an over-

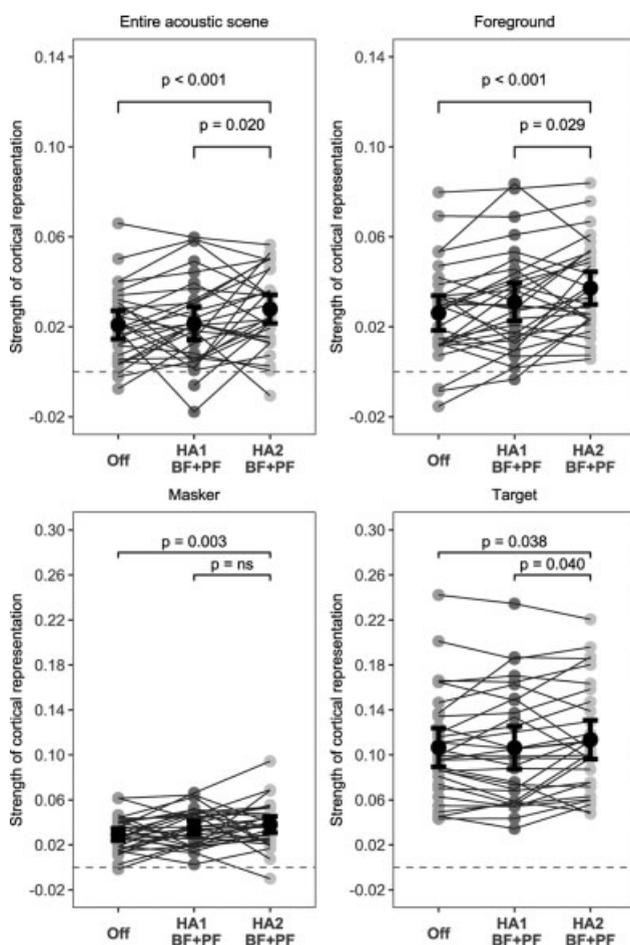
view and differences in participants and test setups are provided here. The reader is referred to the articles mentioned for further methodological details.

Thirty-one experienced hearing-aid users with mild to moderately severe sensorineural hearing loss who qualified for fitting with RITE hearing aids (mean age: 65.6 years) were enrolled in the study. All provided informed consent and the experiments were approved by the Science Ethics Committee for the Capital Region of Denmark (journal no. H20028542). As described in the article by Alickovic et al,<sup>44</sup> two continuous speech signals from different talkers were presented at 73 dB SPL from two different loudspeakers in front of the participants ( $\pm 30$ -degree azimuth). Participants were instructed to attend to one of the foreground talkers (the target talker) and to ignore the other (the masker talker). Meanwhile, babble noise at 70 dB SPL was presented from four loudspeakers in the background ( $\pm 100$ - and  $\pm 153$ -degree azimuth), with a mix of 4 talkers in each loudspeaker. The study

was designed to measure the benefit of noise reduction in HA2 and to compare the noise reduction systems of HA1 and HA2, yielding three test conditions: noise reduction deactivated in HA2 (“Off”) and noise reduction activated in HA1 and HA2 (“BF + PF”). For each test condition, the participants listened to 20 trials of 38 seconds each. Both hearing aids were fitted to participants in the same way as described in Section 5.2.

During this task, EEG was recorded, from which a measure was derived that indicates how strongly parts of the acoustic scene or single sound sources are represented in the auditory cortex.<sup>42,44,46</sup> This measure is referred to as cortical representation. By analyzing the EEG

data in different time windows (see Fig. 3 in Alickovic et al<sup>44</sup>), these cortical representations at different stages of auditory cortical processing can be estimated. Early EEG responses (<85 milliseconds) are thought to originate from the primary areas of the auditory cortex and are less influenced by selective attention so that all sounds in the acoustic scene are co-represented. In contrast, late EEG responses (>85 milliseconds) are generated from higher-order, nonprimary cortical areas and show a large effect of selective attention, such that the cortical representation of the target talker is emphasized compared with that of the masker talker and the background.<sup>47–49</sup> Following this premise, the cortical representation of the entire



**Figure 13** Strength of cortical representation of the entire acoustic scene (top left) and of the foreground (top right) as estimated from early EEG responses, and of the target talker (bottom left) and of the masker talker (bottom right) as estimated from late EEG responses. Gray dots indicate trial-averaged individual results, whereas black dots and error bars show the group strengths of cortical representation (grand average  $\pm 1$  between-subject standard error of the mean). Each horizontal line in gray denotes a single participant.

acoustic scene (comprising target talker, masker talker, and background noise) and of the foreground (comprising target and masker talkers) was investigated using early EEG responses, while the cortical representation of the individual foreground talkers (target and masker) was investigated using late EEG responses.

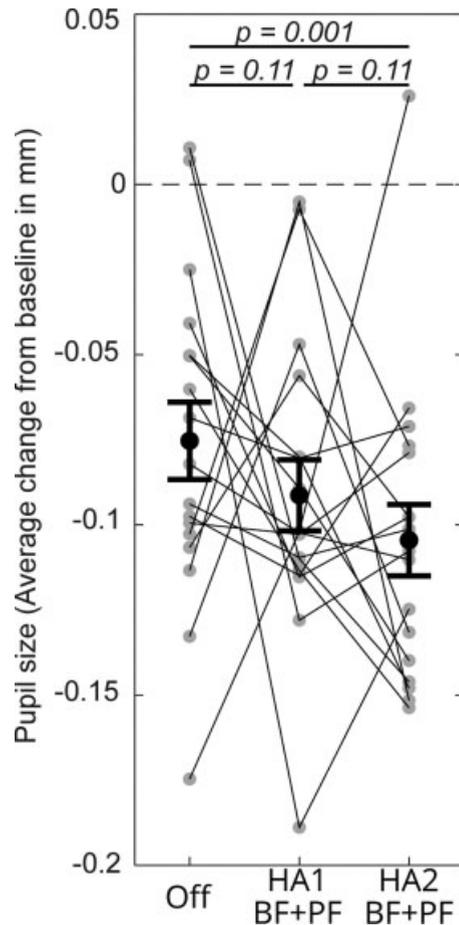
The top panels in Fig. 13 show the strength of the cortical representation of the entire acoustic scene (i.e., the combination of all objects in the environment) and of the foreground (i.e., the combination of the two possible talkers that the user may attend to) based on early EEG responses. A one-way linear mixed model ANOVA revealed a significant main effect of condition (entire acoustic scene:  $F_{2,1232} = 9.4$ ,  $p < 0.001$ ; foreground:  $F_{2,1230} = 11.3$ ,  $p < 0.001$ ). Post hoc pairwise comparisons (Bonferroni corrected) revealed that the strength of early cortical representations was significantly higher for the “BF + PF” conditions than for the “Off” condition (entire acoustic scene:  $p < 0.001$ ; foreground:  $p < 0.001$ ) and significantly higher for HA2 than for HA1 (entire acoustic scene:  $p = 0.020$ ; foreground:  $p = 0.029$ ). These results suggest that activating noise reduction contributes to a more accurate representation of the hearing aid user’s whole listening environment in the early stages of cortical processing. The same can be said about foreground sound sources that may become the focus of attention. Finally, the results suggest that the DNN-based noise reduction system of HA2 is more advantageous in these regards.

The bottom panels in Fig. 13 show the strength of the cortical representation of the target and masker talkers based on late EEG responses. A one-way linear mixed model ANOVA revealed a significant main effect of condition (target:  $F_{2,1225} = 4.1$ ,  $p = 0.016$ ; masker:  $F_{2,1226} = 5.6$ ,  $p = 0.004$ ). Post hoc pairwise comparisons (Bonferroni corrected) showed that the strength of late cortical representations was significantly higher for “BF + PF” conditions than for the “Off” condition (target:  $p = 0.038$ ; masker:  $p = 0.003$ ) and significantly higher for HA2 than for HA1 for the target talker ( $p = 0.040$ ). These results suggest that the tested noise reduction systems help the user selectively attend to a talker of interest in complex listening environments while maintaining access to other secondary talkers, which is important to

allow the user to switch attention as the situation calls for it. The DNN-based HA2 seems to provide a greater advantage in this regard.

Finally, the pupil size of 17 of the participants was recorded while they selectively attended to the target talker during the same EEG experiment. Pupil size indicates how much cognitive effort is spent on a listening task.<sup>45,50,51</sup> As a general rule, a smaller pupil size indicates reduced listening effort compared with a larger pupil size.

The pupillometry results (Fig. 14) showed a significant difference between test conditions (one-way ANOVA,  $F_{2,937} = 5.3$ ,  $p = 0.005$ ). Post hoc tests revealed that there was a significantly smaller pupil size for HA2 “BF + PF” compared with “Off” ( $t_{931} = -3.2$ ,  $p = 0.001$ ),



**Figure 14** Pupil size depicted as the average change from baseline. Black dots and error bars indicate the average across participants (mean  $\pm$  1 between-subject standard error of the mean). Gray dots and lines depict individual means across trials.

while the other two comparisons did not reach significance (HA1 “BF + PF” vs. “Off”:  $t_{931} = -1.6$ ,  $p = 0.11$ ; HA2 “BF + PF” vs. HA1 “BF + PF”:  $t_{931} = -1.6$ ,  $p = 0.11$ ). This indicates that the noise reduction system of HA2 reduces listening effort during a selective-attention task in a complex multitalker noisy environment, in line with the findings of Fiedler et al.<sup>45</sup>

In summary, the studies discussed here indicate that noise reduction systems in commercial hearing aids which combine an MVDR beamformer with a postfilter can provide clinical benefits to users, with the most significant effects obtained with the DNN-based HA2. Benefits are seen in terms of increased speech intelligibility in noise, stronger cortical representations of multiple sound sources, and reduced listening effort.

## 5: CONCLUSION

Noise reduction in modern hearing aids typically takes the form of joint beamforming and post-filtering, which work particularly well when the noise is separate from the target speech in either time, frequency, or direction of arrival. Rapid advances in machine learning are increasingly influencing the design approach to such systems. In fact, hearing aids using neural networks for postfiltering are already commercially available.

Experimental results presented in this article indicate that noise reduction algorithms provide a range of benefits. First, they can improve SNR and speech intelligibility in noisy environments. Second, they can decrease listening effort and improve the user’s ability to focus on specific targets. As discussed here, improvements in noise reduction algorithms are highly relevant because they effectively extend the range of listening environments in which hearing aids can benefit the user.

## CONFLICT OF INTEREST

None declared.

## ACKNOWLEDGMENTS

The authors would like to thank Micha Lundbeck (HörTech gGmbH) and Michael Schulte (Hörzentrum Oldenburg GmbH) for their

contributions to data collection and analysis related to output SNR and speech intelligibility measurements, as well as the following colleagues from Oticon and Eriksholm Research Centre for their contributions to the research studies reported in this manuscript: Josefine Juul Jensen, Carina Graversen, Dorothea Wendt, Elaine Hoi Ning Ng, Hamish Innes-Brown, Brian Kai Loong Man, Sara Al-Ward, and Louis Villejouberts. Lastly, the authors would like to thank Joshua M. Alexander whose inputs greatly improved this article.

## REFERENCES

1. Kochkin S. MarkeTrak VIII: consumer satisfaction with hearing aids is slowly increasing. *Hear J* 2010; 63(01):19–32
2. Picou EM. MarkeTrak 10 (MT10) survey results demonstrate high satisfaction with and benefits from hearing aids. *Semin Hear* 2020;41(01): 21–36
3. Moore BCJ. *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues*. 2nd ed. Wiley; 2007
4. Plomp R. Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J Acoust Soc Am* 1978;63(02):533–549
5. Lopez RS, Bianchi F, Fereczkowski M, Santurette S, Dau T. Data-driven approach for auditory profiling. In: *Proceedings of the International Symposium on Auditory and Audiological Research*. Nyborg, Denmark 2017:247–254
6. Moore B. *An Introduction to the Psychology of Hearing*. 6th ed. Leiden, Netherlands: Brill; 2013
7. Elko GW. Superdirectional microphone arrays. In: *Gay SL, Benesty Jeds.. Acoustic Signal Processing for Telecommunication*. New York, United States: Springer; 2000:181–237
8. Capon J. High-resolution frequency-wavenumber spectrum analysis. *Proc IEEE* 1969;57(08): 1408–1418
9. Cox H, Zeskind R, Owen M. Robust adaptive beamforming. *IEEE Trans Acoust Speech Signal Process* 1987;35(10):1365–1376
10. Wiener N. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications*. Cambridge: MIT Press; 1949
11. Ephraim Y, Malah D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Trans Acoust Speech Signal Process* 1984;32(06):1109–1121
12. Gannot S, Vincent E, Markovich-Golan S, Ozerov A. A consolidated perspective on multimicrophone speech enhancement and source separation. *IEEE/*

- ACM Trans Acoust Speech Signal Process 2017;25(04):692–730
13. Kjems U, Boldt JB, Pedersen MS, Lunner T, Wang D. Role of mask pattern in intelligibility of ideal binary-masked noisy speech. *J Acoust Soc Am* 2009;126(03):1415–1426
  14. Doclo S. Multi-microphone noise reduction and dereverberation techniques for speech applications. PhD thesis KU Leuven Leuven, Belgium 2003
  15. Simmer KU, Bitzer J, Marro C. Post-filtering techniques. In: Brandstein M, Ward Deds. *Microphone Arrays: Signal Processing Techniques and Applications*. Springer; 2001:39–60
  16. Kjems U, Jensen J. Maximum likelihood based noise covariance matrix estimation for multi-microphone speech enhancement. In: *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. Bucharest, Romania 2012:95–299
  17. Jensen J, Pedersen MS. Analysis of beamformer directed single-channel noise reduction system for hearing aid applications. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* South Brisbane, Queensland, Australia 2015:5728–5732
  18. Boldt J, Kjems U, Pedersen MS, Lunner T, Wang D. Estimation of the ideal binary mask using directional systems. In: *Proceedings of the 11th International Workshop on Acoustic Echo and Noise Control Seattle, Washington USA 2008*
  19. Neher T, Wagener KC. Investigating differences in preferred noise reduction strength among hearing aid users. *Trends Hear* 2016;20:20
  20. Kim G, Loizou PC. Gain-induced speech distortions and the absence of intelligibility benefit with existing noise-reduction algorithms. *J Acoust Soc Am* 2011;130(03):1581–1596
  21. Dillon H. *Hearing Aids*. Thieme; 2000
  22. Kuklasinski A, Doclo S, Jensen SH, Jensen J. Maximum likelihood PSD estimation for speech enhancement in reverberation and noise. *IEEE/ACM Trans Acoust Speech Signal Process* 2016;24(09):1599–1612
  23. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press; 2016
  24. Bishop C. *Pattern Recognition and Machine Learning*. Springer; 2006
  25. Brown T, Mann B, Ryder N et al. Language models are few-shot learners. In: Larochelle H, Ranzato M, Hadsell R, Balcan MF, Lin Heds. *Advances in Neural Information Processing Systems*. 2020;33:1877–1901
  26. Kolbæk M, Tan ZH, Jensen J. Speech intelligibility potential of general and specialized deep neural network based speech enhancement systems. *IEEE/ACM Trans Acoust Speech Signal Process* 2017;25(01):153–167
  27. Healy EW, Yoho SE, Wang Y, Wang D. An algorithm to improve speech recognition in noise for hearing-impaired listeners. *J Acoust Soc Am* 2013;134(04):3029–3038
  28. Healy EW, Yoho SE, Chen J, Wang Y, Wang D. An algorithm to increase speech intelligibility for hearing-impaired listeners in novel segments of the same noise type. *J Acoust Soc Am* 2015;138(03):1660–1669
  29. Chen J, Wang Y, Yoho SE, Wang D, Healy EW. Large-scale training to increase speech intelligibility for hearing-impaired listeners in novel noises. *J Acoust Soc Am* 2016;139(05):2604–2612
  30. Kim G, Lu Y, Hu Y, Loizou PC. An algorithm that improves speech intelligibility in noise for normal-hearing listeners. *J Acoust Soc Am* 2009;126(03):1486–1494
  31. Xu Y, Du J, Dai L, Lee C. An experimental study on speech enhancement based on deep neural networks. *IEEE Signal Process Lett* 2014;21(01):65–68
  32. Minnaar P, Albeck SF, Simonsen CS, Søndersted B, Oakley SAD, Bennedbæk J. Reproducing real-life listening situations in the laboratory for testing hearing aids. In: *Audio Engineering Society Convention 135: Paper 8951*, 2013
  33. Kirkeby O, Nelson PA, Hamada H, Orduna-Bustamante F. Fast deconvolution of multichannel systems using regularization. *IEEE Trans Acoust Speech Signal Process* 1998;6(02):189–194
  34. Brumm H, Zollinger S. The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour* 2011;148(11/13):1173–1198
  35. Berouti M, Schwartz R, Makhoul J. Enhancement of speech corrupted by acoustic noise. *IEEE Trans Acoust Speech Signal Process* 1979:208–211
  36. Andersen AH, Jensen J, Pedersen M et al. Hearing device comprising a noise reduction system. *United States Patent Application Publication No. US 2020/0260198 A1*
  37. Hagerman B, Olofsson Å. A method to measure the effect of noise reduction algorithms using simultaneous speech and noise. *Acta Acust United Acust* 2004;90(02):356–361
  38. Wardenga N, Batsoulis C, Wagener KC, Brand T, Lenarz T, Maier H. Do you hear the noise? The German matrix sentence test with a fixed noise level in subjects with normal hearing and hearing impairment. *Int J Audiol* 2015;54(Suppl 2):71–79
  39. Buus S, Florentine M. Growth of loudness in listeners with cochlear hearing losses: recruitment reconsidered. *J Assoc Res Otolaryngol* 2002;3(02):120–139
  40. Ohlenforst B, Zekveld AA, Jansma EP et al. Effects of hearing impairment and hearing aid amplification on listening effort: a systematic review. *Ear Hear* 2017;38(03):267–281
  41. Ohlenforst B, Wendt D, Kramer SE, Naylor G, Zekveld AA, Lunner T. Impact of SNR, masker type and noise reduction processing on sentence

- recognition performance and listening effort as indicated by the pupil dilation response. *Hear Res* 2018;365:90–99
42. Alickovic E, Lunner T, Wendt Det al.. Neural representation enhanced for speech and reduced for background noise with a hearing aid noise reduction scheme during a selective attention task. *Front Neurosci* 2020;14:846
  43. Lunner T, Alickovic E, Graverson C, Ng EHN, Wendt D, Keidser G. three new outcome measures that tap into cognitive processes required for real-life communication. *Ear Hear* 2020;41(Suppl 1):39S–47S
  44. Alickovic E, Ng EHN, Fiedler L, Santurette S, Innes-Brown H, Graverson C. Effects of hearing aid noise reduction on early and late cortical representations of competing talkers in noise. *Front Neurosci* 2021;15:636060
  45. Fiedler L, Seifi Ala T, Graverson C, Alickovic E, Lunner T, Wendt D. Hearing Aid Noise Reduction Lowers the Sustained Listening Effort During Continuous Speech in Noise—A Combined Pupilometry and EEG Study. *Ear and Hearing*; in press. Doi: 10.1097/AUD.0000000000001050
  46. Fiedler L, Wöstmann M, Herbst SK, Obleser J. Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *Neuroimage* 2019;186:33–42
  47. O’Sullivan J, Herrero J, Smith Eet al.. Hierarchical encoding of attended auditory objects in multi-talker speech perception. *Neuron* 2019;104(06):1195–1209.e3
  48. Zion Golumbic EM, Ding N, Bickel Set al.. Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. *Neuron* 2013;77(05):980–991
  49. Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 2012;485(7397):233–236
  50. Seifi Ala T, Graverson C, Wendt D, Alickovic E, Whitmer WM, Lunner T. An exploratory study of EEG alpha oscillation and pupil dilation in hearing-aid users during effortful listening to continuous speech. *PLoS One* 2020;15(07):e0235782
  51. Pichora-Fuller MK, Kramer SE, Eckert MAet al.. Hearing impairment and cognitive energy: the framework for understanding effortful listening (FUEL). *Ear Hear* 2016;37(Suppl 1):5S–27S