Enhancing Directionality with Deep Neural Networks

AUDIBEL

Dr.-Ing. Daniel Marquardt, Jinjun Xiao, Ph.D., Al Ganeshkumar, B.Sc., Jingjing Xu, Ph.D., Larissa Taylor, Ph.D., Martin McKinney, Ph.D.

Key takeaways:

- Audibel's Aris AI is the world's first multi-channel deep neural network (DNN)-powered directionality system for hearing aids, designed to enhance speech intelligibility while preserving spatial awareness.
- It features two DNN models that intelligently adapt to complex acoustic environments, outperforming traditional systems in both speech clarity and subjective user preference.
- Clinical evaluations show significant improvements in speech reception thresholds and audibility of environmental sounds, especially for off-axis sources, enabling users to engage more naturally with their environments

Introduction

Modern hearing aids leverage advanced signal processing techniques to help individuals with hearing loss navigate complex and dynamic listening environments. Among these techniques, directionality algorithms are critical for enhancing speech understanding in noisy settings by focusing on sounds originating from specific directions – most often in front of the listener. While this directional focus improves frontal clarity in noisy situations, maintaining awareness of sounds from all directions is equally important. Broader auditory awareness supports both social engagement and safety, enabling users to participate in group discussions and navigate their environment with confidence.

Striking the right balance between directional focus and spatial awareness is a nuanced challenge. Overemphasizing directionality can diminish important spatial cues, leading to an unnatural and potentially disorienting listening experience where listeners may struggle to localize sounds. Conversely, insufficient directionality can leave listeners more exposed to background noise, increasing reliance on noise reduction algorithms. While these features can help suppress unwanted sounds, they may also compromise sound quality or inadvertently mask relevant auditory information. Achieving the optimal balance is essential for delivering both clarity and comfort, allowing listeners to engage naturally in conversations while remaining attuned to their surroundings.

Audibel's intelligent directionality system addresses this challenge by integrating a wide range of contextual information about the listener and their environment to apply the most appropriate directional settings without compromising the overall listening experience *(Figure 1)*.

With Audibel's Aris AI, a major advancement in the industry is achieved with the introduction of the world's first multi-channel Deep Neural Network (DNN) directionality algorithm for hearing aids.

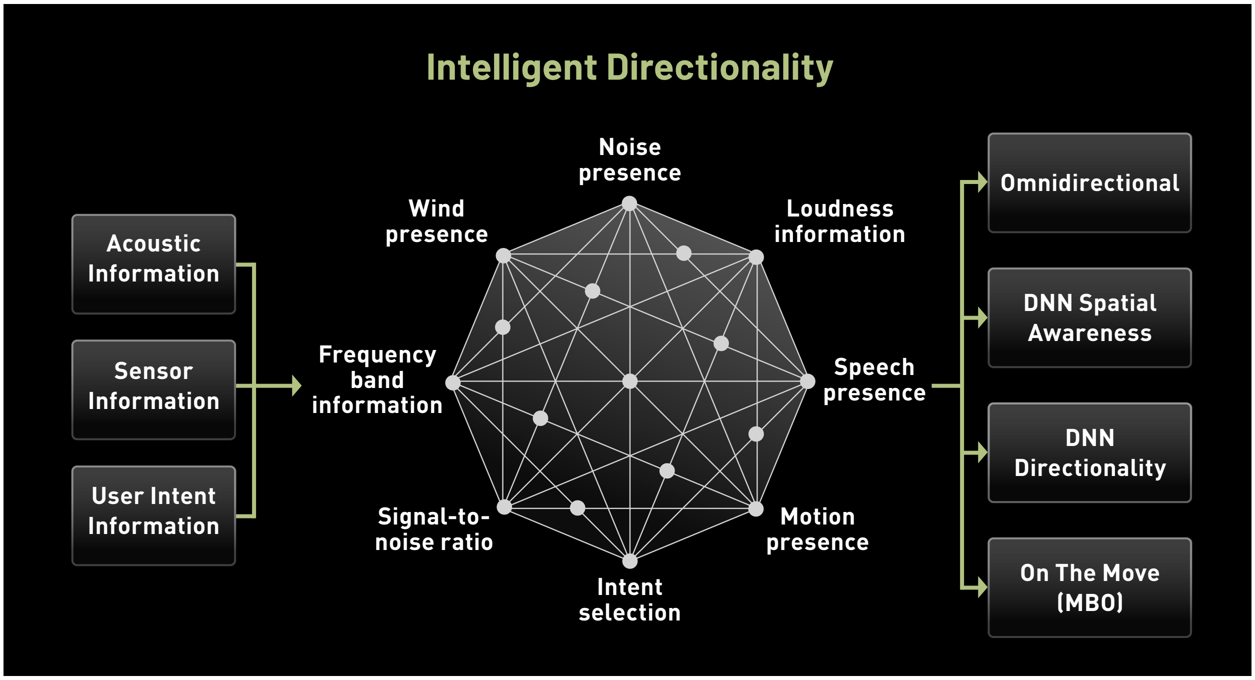

Figure 1: A schematic representation of Audibel's intelligent directionality system.

This system is trained to deliver superior speech intelligibility in noisy environments while preserving the user's spatial awareness. By dynamically adapting to complex listening scenarios, Aris AI offers a more natural and effective listening experience than traditional directional systems – marking a significant step forward in hearing aid technology and quality of life improvement for patients.

At the core of Aris AI are two DNN-powered models that work together to determine when and how to apply directionality. The first model, trained on a large and diverse dataset, predicts the most effective directional parameters for the listener's environment. The second model is designed to detect the subtle environmental cues that indicate when a conversation is occurring. Traditional directionality systems rely on real-time signal statistics from hearing aid microphones, but this approach is vulnerable to fluctuations – even in static environments – due to the time-varying nature of the sound and the need for rapid adaptation. These fluctuations can lead to inconsistent or suboptimal performance.

In contrast, Aris AI's proprietary training and DNN architecture enable it to instantly and reliably predict optimal settings, resulting in more stable and accurate adaptation compared to conventional methods *(Figure 2)*.

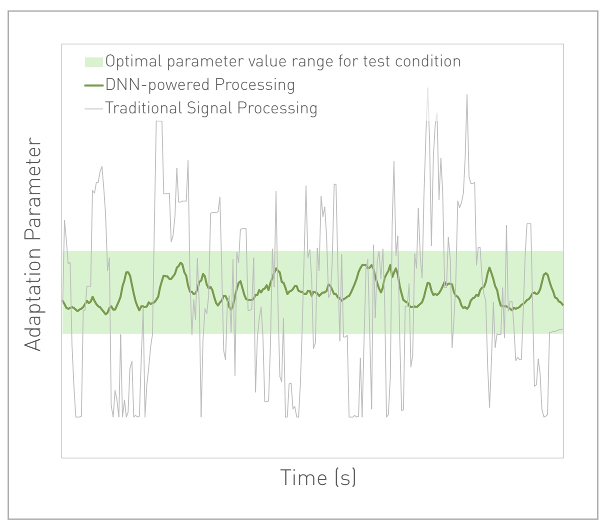

Figure 2: Example depiction of the differences between traditional estimation methods (gray line) versus a DNN-powered model prediction approach (green line) in a laboratory measurement using Aris AI. The figure demonstrates the advantages of the DNN approach in its ability to better predict the most optimal parameters (shaded green area) for a given situation.

When combined with Audibel's existing directionality system, the innovation delivers a highly adaptive and context-aware listening experience across a wide range of real-world scenarios. For instance,

- **In quiet environments,** Aris AI enables a fully omnidirectional mode, preserving both sound quality and spatial awareness for natural perception of speech and the environment.
- **In complex listening environments filled with background noise and overlapping speech,** Aris AI enhances the audibility of important sounds – such as a voice calling out from behind or other environmental sounds like a doorbell – that would typically be suppressed by traditional directionality systems. It intelligently detects when a conversation becomes relevant and adjusts settings to maximize the signal-to-noise ratio (SNR) during speech.
- **When the listener is in motion,** Aris AI supports greater environmental awareness, helping listeners move through their surroundings with increased confidence and safety.

Aris AI Directionality System Evaluation

To evaluate the performance of the Aris AI DNN-powered directionality system, a comparative study was carried out against a legacy directionality feature. The assessment involved a subjective listening test with hearing-impaired participants using the Hearing in Noise Test (HINT). It utilized a paired-comparison test for preference and perceived intelligibility.

Thirteen experienced hearing aid users (6 males, 7 females), aged between 71 and 86 years (mean age 77, with a standard deviation of ±5), with symmetrical mild to moderately severe hearing loss *(Figure 3)*, participated in a laboratory study. Participants were fitted with receiver-in-the-canal (RIC) or micro RIC devices using occluded Power Dome couplings.

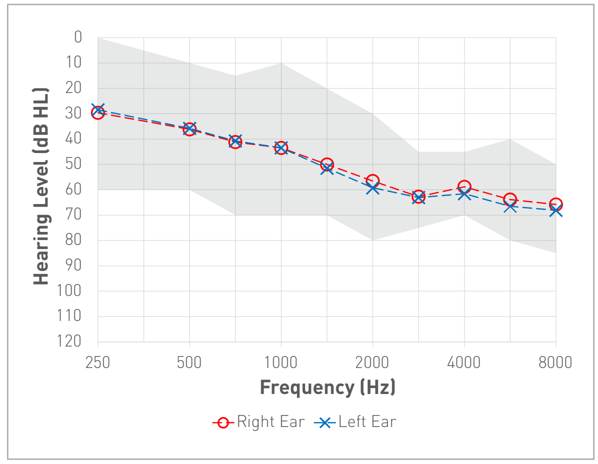

Figure 3: Mean right (red) and left (blue) audiogram and range (shaded gray) of the participants' hearing thresholds (N=13).

Speech Intelligibility Improvements

Test Scenario 1: To demonstrate the newly developed directionality system's ability to improve speech intelligibility for non-frontal target sources, a realistic diffuse babble noise environment was generated using eight loudspeakers arranged in a circular configuration at 45° intervals, each positioned 1 meter from the center. The target speech source was placed at 90° *(see Figure 4[a]*, simulating real-world scenarios such as group conversations, walking in crowded areas, or public transportation.

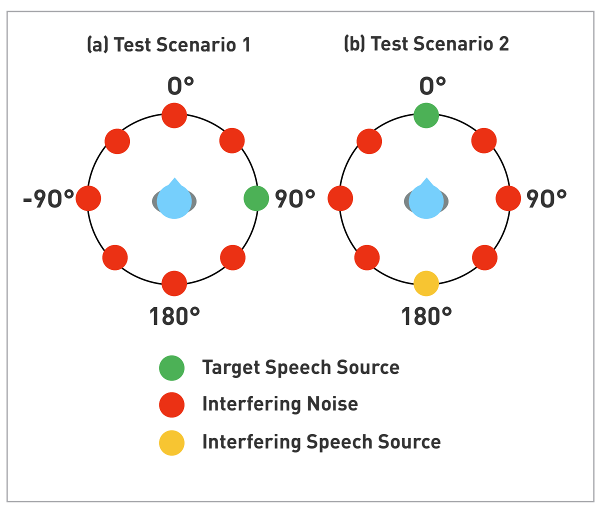

Figure 4: (a) Test Scenario 1 laboratory setup with the target speech source at 90°, and (b) Test Scenario 2 laboratory setup with the target speech source at 0° and an interfering speech source at 180°.

Test Scenario 2: Traditional approaches to preserving speech from non-frontal directions involves switching to an omnidirectional microphone pattern based on a direction-dependent speech-like signal detection mechanism or through user input. However, this approach carries the risk of failing to suppress unwanted interfering signals, such as background conversations or competing speech, which can intrude into the target conversation – particularly in challenging environments like crowded restaurants. To demonstrate that Aris AI's solution offers a more advanced and nuanced approach than simply switching to an omnidirectional pattern, a second scenario was designed that featured the same babble noise used in test Scenario 1. In this setup, a target speech source was positioned at the front, while an additional interfering speech source was placed at the rear *(see Figure 4[b])*. This configuration allows us to evaluate the system's ability to effectively manage competing speech signals while maintaining focus on the target speech. For both scenarios, the noise level was set to 70 dB SPL.

For test Scenario 2, the interfering speech level was set to 75 dB SPL, and for both scenarios the speech level was adapted according to the HINT procedure to measure the speech reception threshold at which a person can correctly recognize and repeat 50% of the words heard (SRT50). For test Scenario 1, we compare the performance of the Aris AI directionality system to a legacy directionality system, while in the test Scenario 2, we compare the Aris AI directionality system to an omnidirectional pattern.

Results: As shown in Figure 5(a), the Aris AI directionality system achieves mean and median SRT50 improvements of 1.4 dB (standard deviation ±2.0) and 2.0 dB (interquartile range 3.3) respectively, over a legacy directionality system. These differences show statistical significance via two-tailed Student's t-test ($p < 0.05$).

For the second scenario, as demonstrated in Figure 4(b), the Aris AI directionality system achieves mean and median SRT50 improvements of 4.7 dB (standard deviation ±2.5) and 5.5 dB (interquartile range 4.0), respectively, over an omnidirectional pattern. These differences also show statistical significance using the two-tailed Student's t-test ($p < 0.05$).

Using the established HINT psychometric function slope of 14 %/dB for hearing-impaired listeners *(Harianawala, J., et al., 2019)*, the SRT50 improvements for test S cenario 1 and 2 were translated to speech intelligibility improvements, which are **19.6% (Mean)** and **28% (Median)** for the first test scenario and **65.8% (Mean)** and **77% (Median)** for the second test scenario.

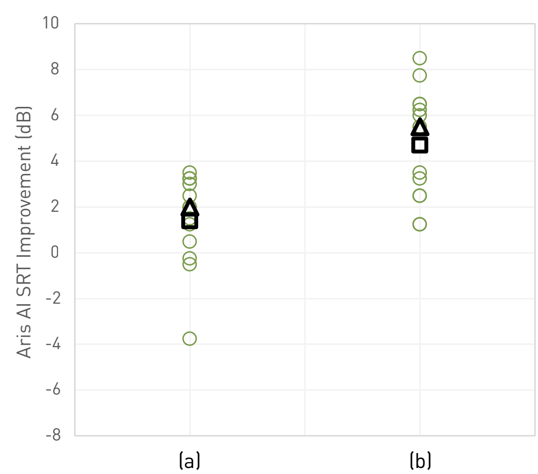

Figure 5: Intelligibility results for the Aris AI directionality feature compared with (a) legacy directionality feature for test Scenario 1 and with (b) omnidirectional pattern for test Scenario 2. The green circles represent individual responses while the black, bold triangles (Δ) and squares (☐) represent the mean and median values respectively. More positive values denote better performance by the Aris AI directionality feature.

This shows that the Aris AI directionality system substantially improves speech intelligibility for non-frontal target sources using an intelligent steering design and is superior to switching to an omnidirectional pattern as demonstrated in the remarkable intelligibility improvements shown for test Scenario 2.

Subjective Preference and Perceived Intelligibility

While SRT50 measurements are essential to demonstrate the speech intelligibility benefits of the new DNN-powered directionality system, it is equally vital to incorporate complementary subjective assessments regarding personal preference and perceived speech intelligibility to make sure that the benefit translates into noticeable improvements, beyond standardized test procedures. For this subjective comparison, the following acoustic scenarios were set up *(see Figure 6[a]–[b])*.

a. S0ND: An 8-speaker diffuse bar noise signal with a female speech signal at 0° (front)
b. S90ND: An 8-speaker diffuse city traffic noise with a female speech signal at 90° (side)
c. S180ND: An 8-speaker diffuse shopping mall noise with a female speech at 180° (rear)
d. S0ND180: An 8-speaker diffuse restaurant noise with female speech at 0° (front) and an interfering speech source at 180° (rear)

Audio recordings were obtained via a manikin with both the Aris AI and a legacy directional setup, with processing parameters calibrated to reflect each participant's individual audiometric profiles. The recordings were then played back to the participant via headphones to allow for a paired comparison between the two directionality systems. For each trial, participants were asked one of the following 2 questions:

1. Which of the 2 signals do you prefer to listen to?
2. In which signal has the target speech the best speech clarity? (for which participants could answer "Signal A", "No difference", "Signal B")

In total there were 8 distinct comparisons (2 different questions, 4 acoustic signals), each repeated 3 times in a randomized order, resulting in a total of 24 comparison trials per participant. To derive a robust representative outcome from the 3 repetitions, we employed a majority vote-filtering method.

This approach aggregates repeated responses by selecting the option that received the most votes across the three presentations of each signal. In cases where no option obtained a majority the aggregated vote was "Same".

Figure 7 presents the overall findings from the "preference" and "clarity" evaluations. In the assessment *(Figure 7[a])*,of the 68 times participants had a preference, 94% favored Aris AI. This preference was statistically significant, as confirmed by a linear mixed-effects model ($p < 0.001$, 95% CI [0.3, 0.53]).

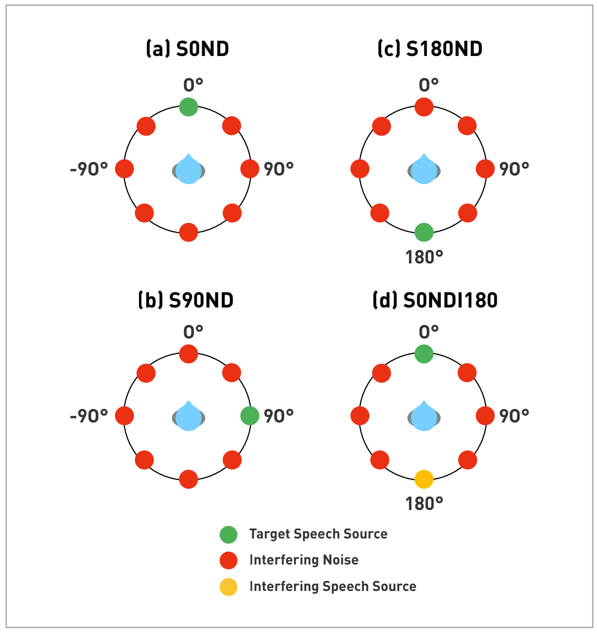

Figure 6: Speech and noise positions for the different acoustic scenarios evaluated.

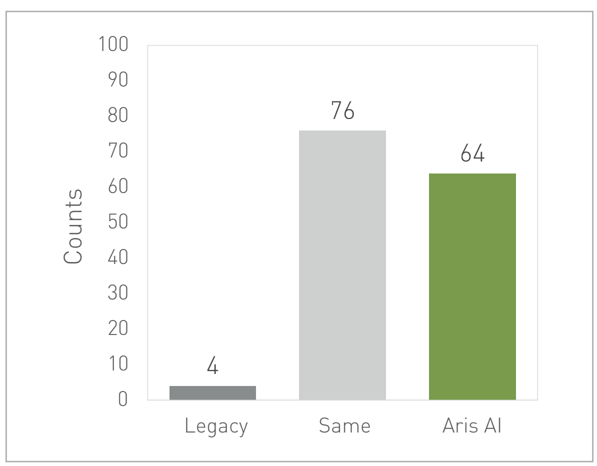

Figure 7(a): Preference for Aris AI directionality vs. legacy directionality across all acoustic scenarios evaluated.

Similarly, for the clarity-related evaluation *(Figure 7[b])*, participants made 83 preference selections, with 96% favoring Aris AI, highlighting its superior directionality system across diverse acoustic environments. These clarity results were also highly significant ($p < 0.001$, 95% CI [0.41, 0.66])

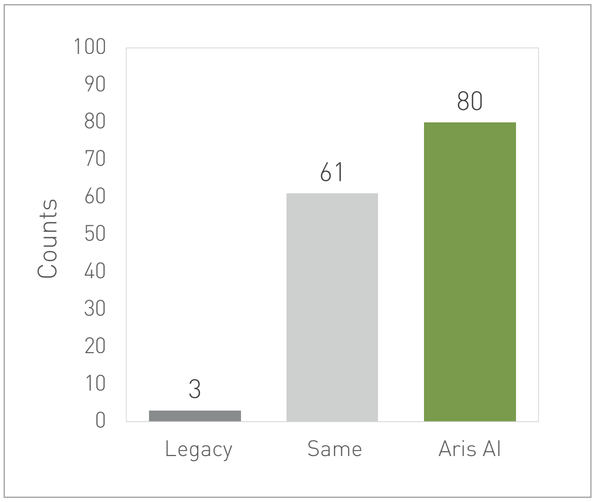

Figure 7(b): Speech clarity for Aris AI directionality vs legacy directionality across all acoustic scenarios evaluated.

Spatial Awareness Evaluation

While the use of directionality is a highly effective method to enhance speech understanding, it becomes less desirable in situations where speech is absent. In such cases, excessive directionality can reduce spatial awareness, particularly for sounds coming from the rear hemisphere. This can make it more difficult for listeners to detect important environmental cues, such as a footstep, doorbells, or a voice from behind.

Nineteen experienced hearing aid users (11 males, 8 females) with symmetrical mild to moderately severe hearing loss *(Figure 8)* took part in this study to evaluate spatial awareness using the Aris AI directionality system. Participants ranged in age from 26 to 82 years (mean age 68.9, standard deviation ±13.8). Of the group, 5 were fitted with open-fit coupling, while the remaining 14 used occluded fittings.

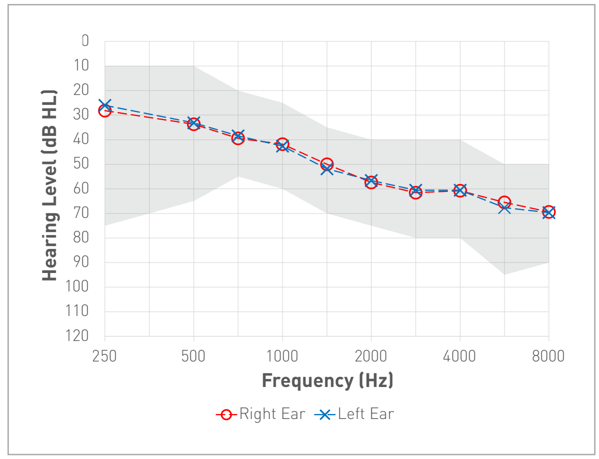

Figure 8: Mean right (red) and left (blue) audiogram and range (shaded gray) of the participants' hearing thresholds (N=19).

Audibility of Environmental Sounds

Participants compared the audibility of environmental sounds with two different settings: a directional setting and another one created specifically to help with environmental awareness (Spatial Adaptation mode). The lab setup was as shown in Figure 9 – the background babble noise was played at 70 dB SPL, with target sounds presented either at 135° or 180°. Participants were instructed to first find the softest audible level for the target sounds and then compare it between the two settings to indicate which was more audible.

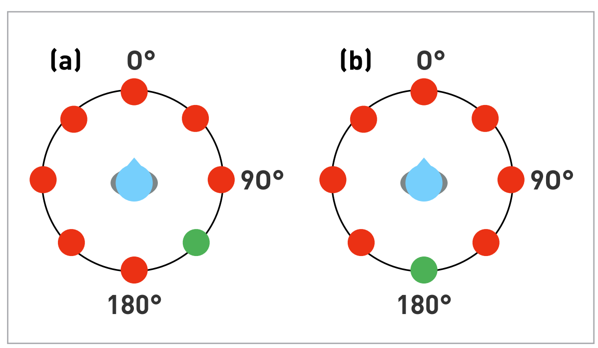

Figure 9: Two test conditions with background noise (red) and target sounds (green) presented at (a) 135° and (b) 180° azimuth.

The results in Figure 10 show that overall, the Spatial Adaptation mode improves audibility by over 60% when compared with the directional mode. For the doorbell sound in particular, the Spatial Adaptation mode improves audibility by over 90% from both test angles evaluated. Although the results showed better audibility for the footstep condition at the 135° angle, the difference did not achieve significance.

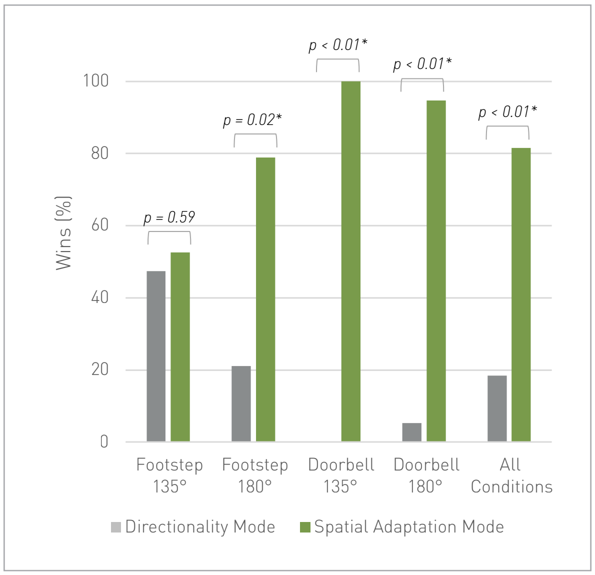

Figure 10: Audibility test results averaged across the two test conditions for the two target sounds. Using the exact binomial test, significant differences were observed (*) in three out of the four conditions, as well as across all conditions tested.

Improved Off-axis Speech Onset

Aris AI was developed to enhance speech signal-to-noise ratio (SNR) from the very beginning of speech and throughout ongoing conversations. As previously mentioned, traditional directional processing tends to attenuate sounds from the rear, which can result in the initial onset of speech from behind being suppressed before adaptive systems have time to respond optimally. This section presents results demonstrating how enabling the Spatial Adaptation mode significantly improves off-axis speech onset SNR, as well as Speech Reception Thresholds (SRT-50).

These improvements allow listeners to better perceive greetings or speech originating from behind, enabling more natural and timely responses.

In a laboratory study, an Aris AI RIC RT device was fitted with an N3 audiogram with a power dome. All hearing aid features were at default settings, except for the feedback canceller, which was disabled to accommodate the Hagerman method (2004) for SNR calculation. Recordings were made in a sound-treated lab using an 8-speaker array and a Brüel & Kjær Head and Torso Simulator manikin. Food court noise from the ARTE database was played at 70 dB SPL from all speakers, with target speech presented at 67 dB SPL from 135 degrees. SNR was measured with and without Binaural Spatial Adaptation enabled. As shown in Figure 11, enabling Binaural Spatial Adaptation consistently improved the SNR for the target speech sounds, with up to an 8dB improvement.

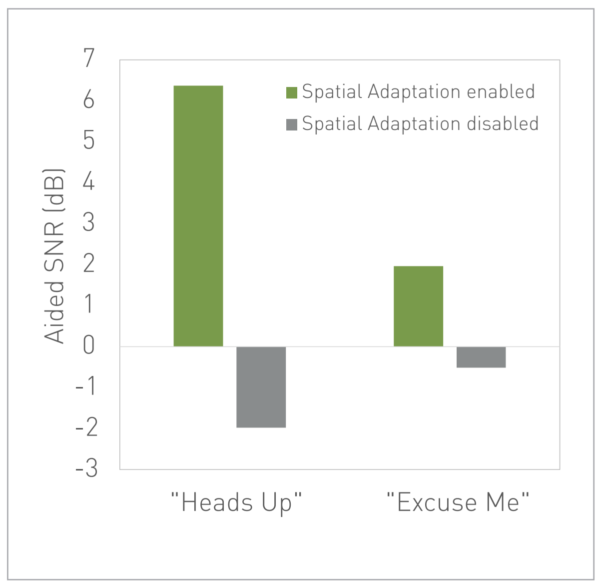

Figure 11: Aided SNR measurement comparing Spatial Adaptation mode (enabled versus disabled) in two speech conditions.

The speech audibility advantage from the rear location was also clinically evaluated using the same participant pool from the previous section. This study aimed to assess the detectability of speech presented from the rear hemisphere (135° and 180° azimuths, see Figure 9) using spondee words.

The goal was to determine the lowest sound level at which participants could correctly repeat approximately 50% of the words. Speech materials consisted of 51 spondee words (e.g. “cowboy”, “birthday”, “cupcake”) sourced from the Auditec CD (The Auditec CD Spondees, 2015). Background noise was fixed at 70 dB SPL, while speech stimuli was initially presented at 76 dB SPL (+6 dB SNR). Participants were instructed to repeat each word aloud. A 4-dB step-down was applied following a correct response, and a 2-dB step-up followed an incorrect response. This adaptive procedure continued until the participant responded correctly at the same level on two consecutive presentations.

The results, summarized in Figure 12, indicate that the Binaural Spatial Adaptation mode significantly improved speech reception thresholds compared to traditional directional settings. Specifically, thresholds were lower by approximately 2 dB when speech originated from the rear (135° and 180°), with mean values of -2.21 dB (SD = 3.11) and -1.58 dB (SD = 2.71) respectively, compared to 0.21 dB (SD = 2.66) and -0.53 dB (SD = 3.26) in directional mode.

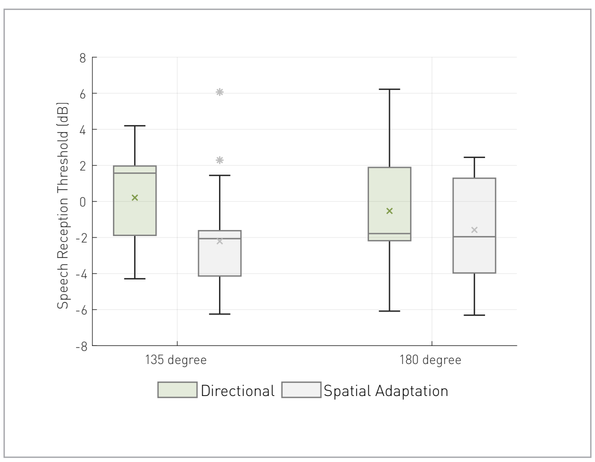

Figure 12: Speech reception threshold measurement comparing Directional mode and Spatial Adaptation mode at 135 and 180 degrees.

These findings support the efficacy of Aris AI’s directionality system, particularly the use of Binaural Spatial Adaptation mode for enhancing the detectability of brief, off-axis speech signals.

Conclusion

In this paper, we introduced a novel multi-channel DNN-powered directionality system designed to enhance speech communication—from the initial onset of speech through to full conversation. While speech understanding remains a key priority for users when speech is present, it is equally important to restore spatial awareness without compromising comfort or safety. The comprehensive Aris AI system, leveraging deep neural networks and motion sensors, delivers an optimized and fully automatic listening experience tailored to the user’s environment and activity.

Clinical and laboratory evaluations confirmed the significant advantages of Aris AI over legacy directional systems. Audibility of environmental sounds improved by over 60% and in some instances as high as 90%, and the signal-to-noise ratio (SNR) for initial off-axis speech improved by up to 8 dB. More importantly for sustained conversation, the system demonstrated an impressive improvement in intelligibility up to 28%. Furthermore, in a subjective evaluation of complex acoustic scenes, Aris AI was strongly favored, earning 94% of preference ratings and 96% of clarity ratings when compared to a legacy system.

By intelligently adapting directionality, Aris AI delivers better hearing in diverse listening environments, helping users converse and navigate their environment with confidence.

Acknowledgements

The authors would like to express their gratitude to Michelle Hicks, Sian Halvorsen, Sarah Iverson, Jamie Hand, and Christine Tan for their thoughtful feedback during the drafting of this manuscript.

References

1. Bisgaard, N., Vlaming, M. S., & Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. Trends in amplification, 14(2), 113–120. https://doi.org/10.1177/1084713810379609
2. Hagerman, B., & Olofsson, A. (2004). A method to measure the effect of noise reduction algorithms using simultaneous speech and noise. Acta Acustica United with Acustica, 90(2), 356-361.
3. Harianawala J, Galster J, Hornsby B. Psychometric Comparison of the Hearing in Noise Test and the American English Matrix Test. J Am Acad Audiol. 2019 Apr;30(4):315-326
4. Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. Journal of the Acoustical Society of America, 95(2), 1085–1099.
5. Radford, Alec; Kim, Jong Wook; Xu, Tao; Brockman, Greg; McLeavey, Christine; Sutskever, Ilya (2022).
6. Robust Speech Recognition via Large-Scale Weak Supervision. (https://arxiv.org/abs/2212.04356)
7. Ricketts, T. A. (2001). Directional hearing aids. Trends in Amplification, 5(4), 139-176.
8. The Auditec CD Spondees (Forms & B) : https://auditec.com/2015/09/30/spondees/

Author Biographies

Daniel Marquardt 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. He currently works as a Principal Signal Processing Engineer at Starkey.

Jinjun Xiao received his M.Sc. degree in mathematics and Ph.D. degree in electrical engineering from the University of Minnesota, Twin Cities, in 2003 and 2006, respectively. He then worked as a postdoctoral fellow at Washington University in St. Louis, Missouri. He has been with Starkey as a Signal Processing Research Engineer since 2010. His research interests include audio signal processing, optimization, and sensor array networks. He has published more than 30 peer-reviewed papers and holds over 10 patents in various research areas.

Al Ganeshkumar is a Principal Audio System Engineer at Starkey Hearing Technologies with over 20 years of experience in the audio industry. After completing research on adaptive noise reduction for hearing aids at the University of Southampton's ISVR, he joined Knowles Electronics in 1992, focusing on hearing aid modeling and algorithms. He later held roles at Shure and Bose, leading projects in professional audio, automotive sound systems, active noise management, and voice pickup technologies. Now back in the hearing aid industry, Al is dedicated to optimizing audio experiences for the hearing impaired. He holds over 50 patents in audio and noise reduction technologies.

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.

Larissa Taylor, Ph.D., is an Audio Systems Engineer at Starkey. Her primary focus is ensuring the overall sound quality of hearing aids. Larissa started at Starkey in 2022 upon completing her doctorate on listening effort and sound quality with hearing aids at McMaster University, Hamilton, Ontario Canada.

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.