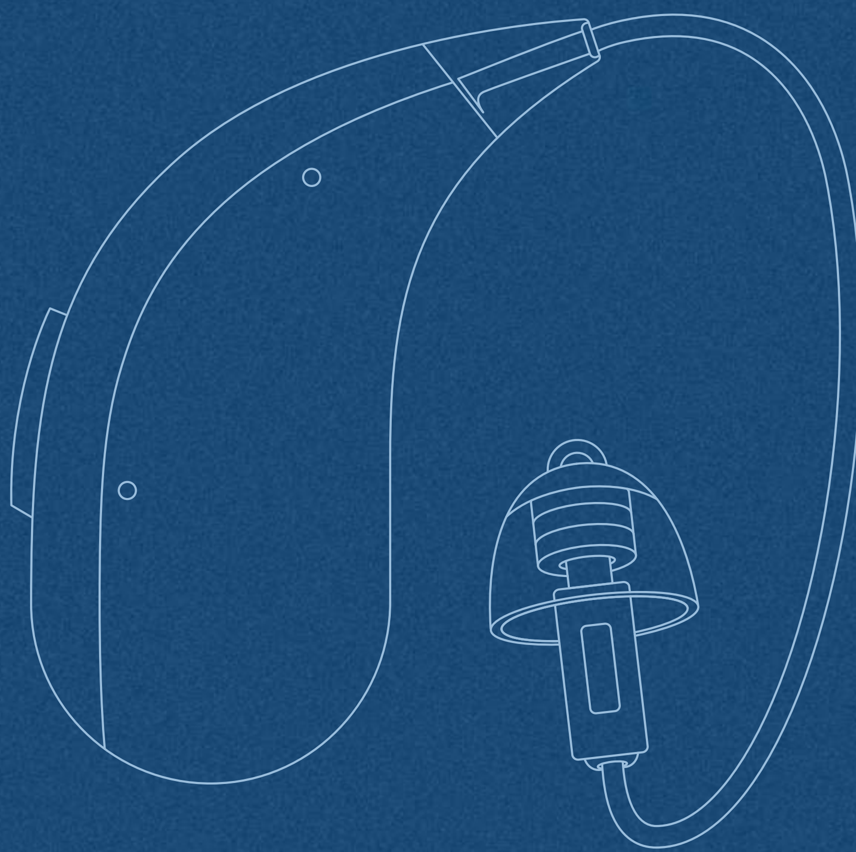
Fortell Research

The Impact of Spatial AI on Speech Intelligibility



VOL. 1

Fortell exists to push the boundaries of hearing technology. Rigorous scientific study underpins our approach and substantiates our progress.

The Problem

Hearing has always been hardest where it matters most

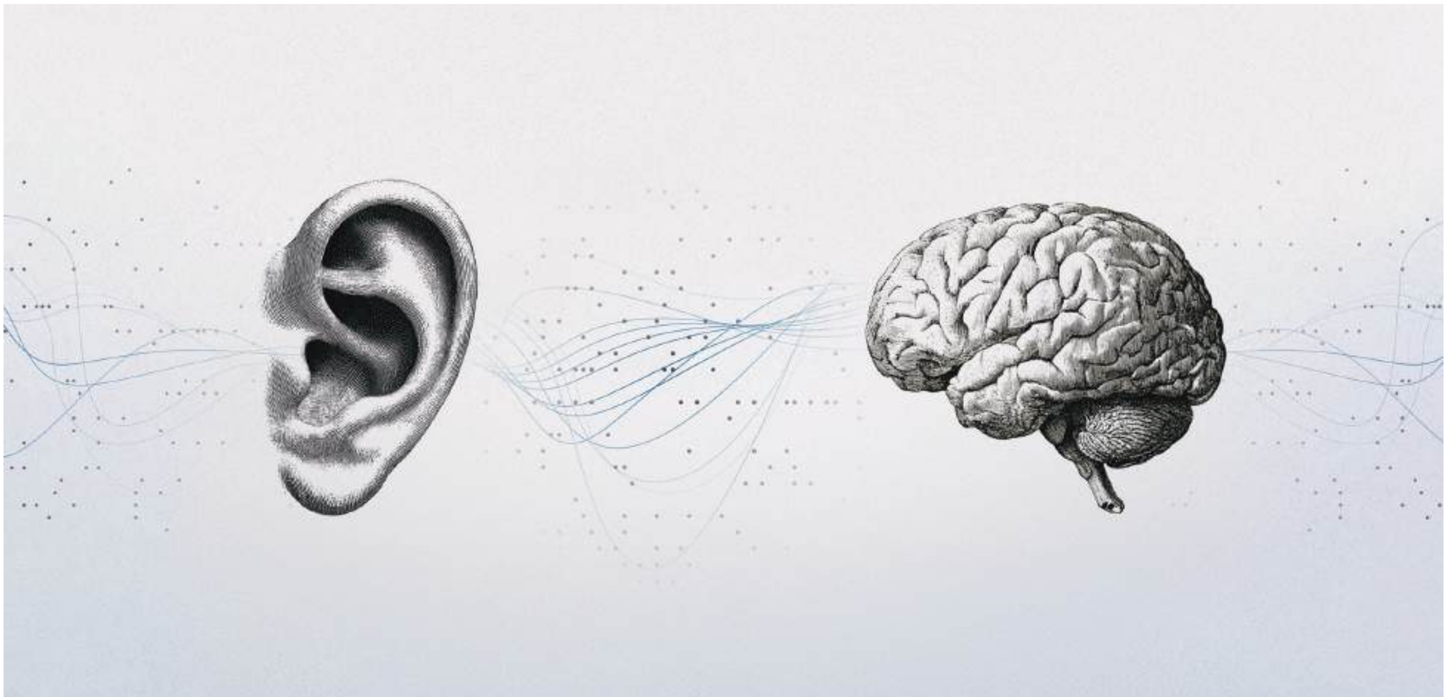
Coffee shops, restaurants, and family dinners—these are the moments when hearing clearly matters. Yet they are also the settings where even the most advanced traditional hearing aids fall short.

Over the past decade, hearing technology has advanced in small steps. Features like Bluetooth, rechargeable batteries, and of late, "AI" have incrementally improved life with hearing loss, but on the core challenge of helping users hear more accurately in noise, none have moved the needle.

> Conventional hearing aids have not solved the longstanding challenge of improving speech understanding in noise.

Research Hypothesis

Hearing aids must assist the brain, not just the ear, to improve speech intelligibility



The human brain is remarkably good at focusing on a single voice in a noisy, crowded environment. It continuously homes in on important talkers, suppresses competing voices, and adapts as conversations shift.

Conventional hearing aids process sound very differently. They're simple: they consume inputs from the world as anonymous streams of frequencies, amplitudes, and time, giving equal credence and amplification to all sounds.

Traditional hearing aids employ techniques like "beamforming" and "noise reduction" but any experienced wearer can speak to the harsh truth: hearing in noise, particularly multi-speaker noise, is a constant struggle.

In recent years, traditional hearing aids have promised artificial intelligence as the next frontier —yet none have materially improved speech understanding in noisy environments for a wide range of hearing aid wearers. Until now.

> Fortell's innovation: a sophisticated source-separating neural network that understands both what a sound is, and where it's coming from, so it can spotlight the voices that matter. We call it Spatial AI.

Technology Solution

Spatial AI:
A new way to process sound

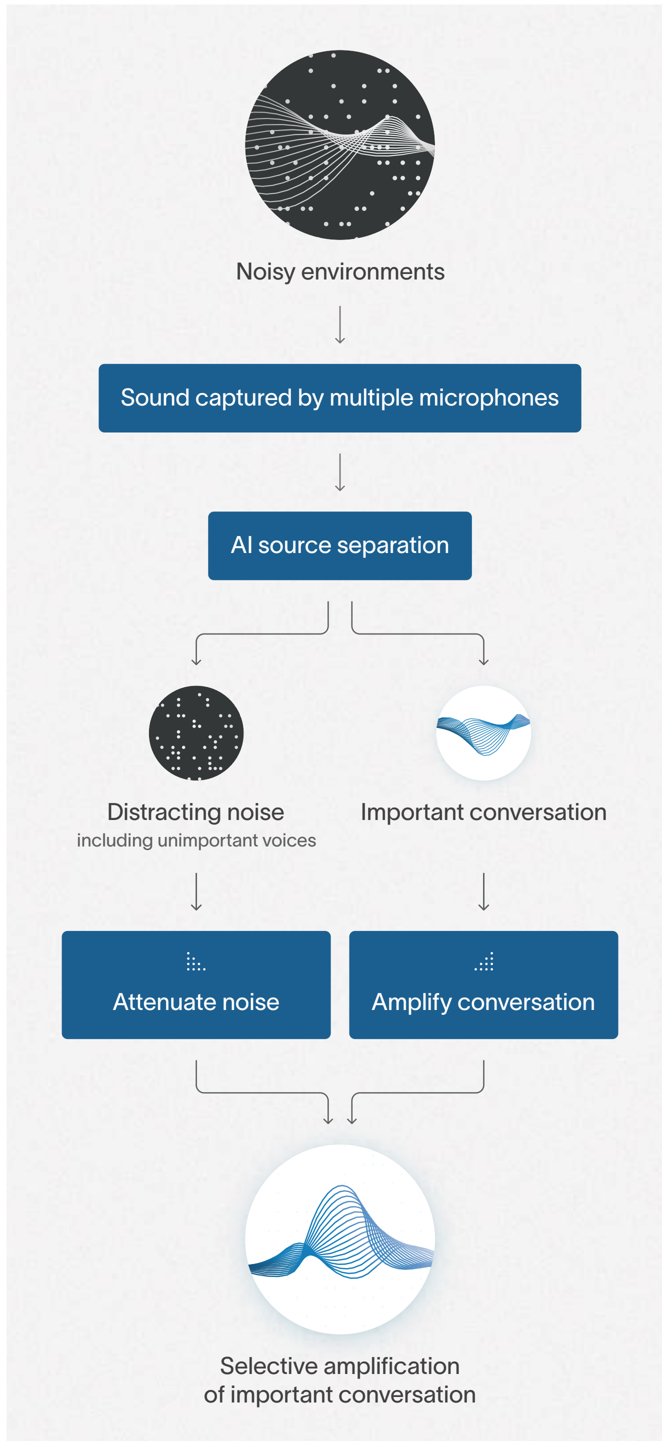
Radical improvement required a radical new technical architecture.

Fortell's Spatial AI is powered by a custom AI processor that consumes sound from multiple microphones to localize voices and isolate them from their surroundings.

Subtle differences in timing, spectral content, and amplitude between multi-microphone inputs offer clues about where a sound is coming from. Fortell's Spatial AI has been trained on millions of real-world audio samples to recognize those clues and lock into the voices that matter, while suppressing the voices that don't.

> This study asks: Does Spatial AI meaningfully improve speech intelligibility compared to the leading AI hearing aid?

To rigorously evaluate its efficacy, Fortell researchers partnered with professors at NYU Langone to design and oversee a study comparing Spatial AI to the market-leading "AI" hearing aid.



Study at a glance

Overview

Participants	30 experienced hearing-aid users with mild to moderately-severe hearing loss
Devices compared	Fortell vs. the leading AI hearing aid
Test design	Double-blind randomized controlled trial
Test environment	Multi-talker, speech-in-noise speaker array designed to mimic challenging, real-world conditions
Primary outcome measure	Participants were asked to repeat words they heard in the presence of background noise. Correct replies were counted and scored
Investigators	Dr. William Shapiro (Director of Audiology, NYU Langone) and Dr. Mario Svirsky (Professor of Hearing Science, NYU Langone) In collaboration with Fortell's licensed audiologists and research scientists
Preregistration	Consistent with scientific best practices, the design, analysis, and methodology for this trial were publicly preregistered with the Open Science Foundation prior to data collection

Trial Design

How the test worked

Participants alternated between Fortell and the control hearing aid in a double-blind protocol, such that neither the participant nor the researchers knew which device was in use at any given time.

Before each testing session, both devices underwent acoustic verification to ensure they were appropriately tuned for each wearer and configured to maximize noise-reduction and directional-focus settings.

During testing, participants listened to sentences presented by a loudspeaker positioned directly in front of them, while competing talkers played from one of three rear loudspeakers.

Participants repeated the words they heard, and each correctly identified word was tabulated to provide an objective measure of speech comprehension. The researcher scoring the results was seated behind a screen and did not know which device the participant was wearing at any given time.

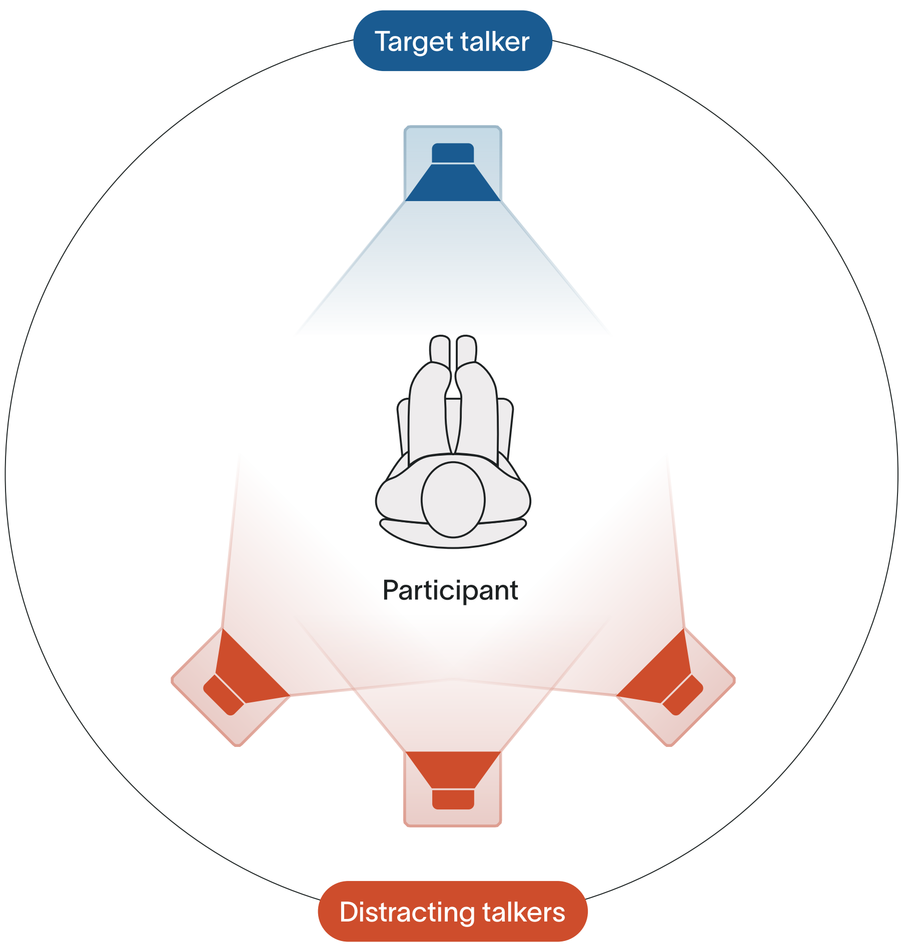


Diagram illustrating position of speech and noise sources relative to the participant during testing. Noise (multi-talker babble) was played out of one of the three loudspeakers behind the participant.

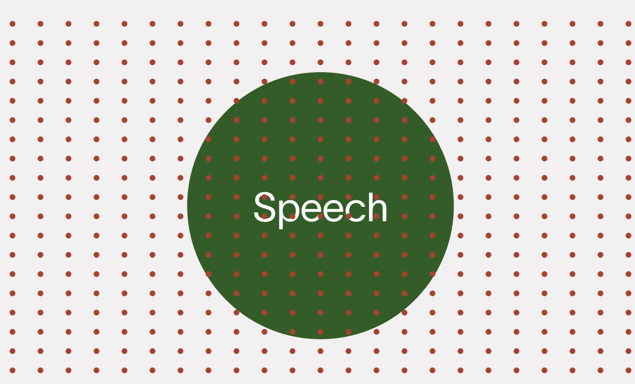
Trial Design

Testing speech understanding across noise levels

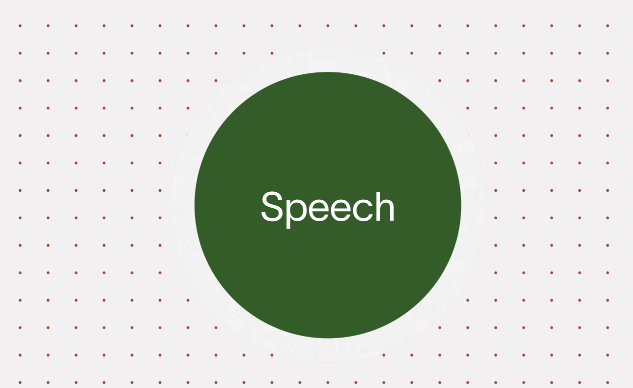
Each participant was tested under three signal-to-noise ratio conditions:

Moderate noise +6 dB	Challenging noise 0 dB	Extremely challenging noise -6 dB
Target speech is ***louder*** than background noise	Target speech and background noise are ***equally*** loud	Target speech is ***quieter*** than the surrounding noise
Equivalent to a quiet café or small meeting	Comparable to a busy restaurant family dinner, or open office	Similar to a crowded bar, train station, or noisy gathering
Following conversation is generally easy for someone with healthy hearing	Most people with healthy hearing can follow conversation with mild concentration	Even healthy hearing listeners may struggle to hear all words correctly

Signal-to-Noise Ratio (SNR) compares the loudness of the sound you want to hear to the loudness of the sounds you don't.



Low SNR – Speech is quieter than background noise; understanding is more difficult



High SNR – Speech is louder than background noise; understanding is easier

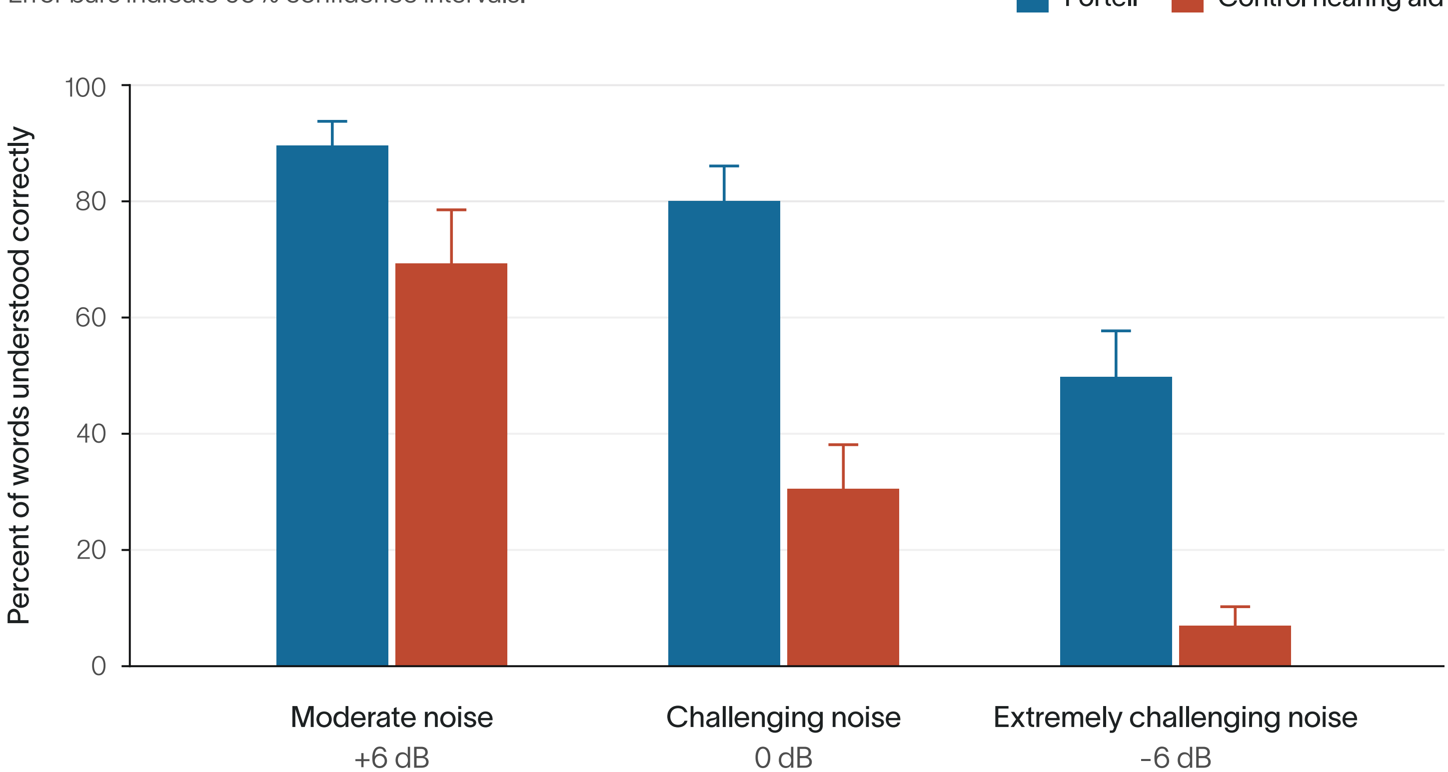
Outcome

The results

Across all participants and all conditions, speech understanding was consistently higher with Fortell compared to the market-leading AI hearing aid. The Fortell advantage widened as the signal-to-noise ratio got more difficult.

Clear gains where hearing usually fails

Error bars indicate 95% confidence intervals.



19x

higher odds of understanding words correctly vs. top-of-the-line AI competitor in the most challenging environments

50x

clearer speech via a 17 dB improvement in signal-to-noise ratio

98%

of background noise eliminated

Read the full paper

Spatial AI Improves Speech Intelligibility for Hearing Aid Wearers in Challenging Multi-talker Noise

Cole Morris[1,4], Igor Lovchinsky[1,4], Christina Callahan[1], Nathan Agmon[1], Kaitlyn Lyngaas[1], Israel Malkin[1], William Shapiro[2], Mario Svirsky[2,3]

[1]Fortell Research
[2]Department of Otolaryngology-HNS, NYU Langone Health
[3]Neuroscience Institute, NYU Langone Health
[4]Co-first authors/contributed equally to this work

Abstract

Objectives: Multi-talker environments represent the most challenging acoustic conditions for individuals with hearing loss. While directional processing and conventional digital noise reduction provide modest benefits, separating target speech from competing speech remains challenging. This study evaluated a new hearing aid that runs a novel spatial neural network-based algorithm (Spatial AI) designed to selectively enhance speech from the front while attenuating both competing talkers (from other directions) and background noise in real-time.

Design: In a pre-registered (available at https://doi.org/10.17605/OSF.IO/SN5HW), double-blind, randomized controlled trial, 30 hearing aid users completed standard speech-in-noise testing (QuickSIN) wearing both an investigational hearing aid (the Fortell AI hearing aids) running Spatial AI and a commercially available control hearing aid employing neural-network-based noise-reduction processing. Both devices were set to maximize available directionality and noise reduction settings. Sentences were presented at three challenging signal-to-noise ratios (SNRs) and the number of words repeated correctly was recorded for each sentence. Speech intelligibility results were compared to lab-measured SNR (using the Hagerman-Olofsson phase-inversion method) and the Hearing-Aid Speech Perception Index (HASPI) v2 metrics.

Results: The investigational hearing aid showed significant improvements in speech intelligibility, the odds of correct word identification increasing 6.0x, 10.6x and 18.9x across SNRs of 6 dB, 0 dB and -6 dB respectively (all $p \ll 0.0001$). The SNR-50 (the SNR required for 50% intelligibility) improved from 3.1 dB with the control hearing aid to -6.2 dB with the investigational hearing aid, resulting in an intelligibility improvement of 9.2 dB. All 30 participants showed significant improvement with the investigational device, indicating consistent benefit across the study cohort. The lab measurements showed that the SNR (over the unaided condition) improved from 2.3 dB with the control device to 15.5 dB with the investigational hearing aid. HASPI v2 showed a similar level of SNR improvement (15.4 dB).

Conclusions: These findings demonstrate that spatially guided neural-network processing can yield substantial intelligibility benefits in a controlled multitalker configuration where the target speech is frontal and competing speech originates from the rear. The effect sizes in this study greatly exceeded established thresholds for clinical significance, suggesting that the tested algorithm may play a meaningful role in addressing the cocktail party problem for hearing-impaired listeners.

Introduction

Conversation in noisy environments is one of the most challenging situations for individuals with hearing loss. This is because individuals with hearing loss do not just require louder speech, they need a more favorable signal-to-noise ratio (SNR) than individuals with normal hearing to achieve the same level of understanding. This phenomenon is measurable in a clinical environment and referred to as "SNR-loss" (Killion et al. 2004).

Hearing aids use a number of signal processing techniques to improve performance in noisy environments. These techniques are often grouped into two broad categories:

1) Directional processing, such as beamforming, exploits spatial differences between microphones to emphasize sounds arriving from certain directions, typically the front.
2) Digital noise reduction techniques, in contrast, aim to suppress background noise by estimating and attenuating signal components that are unlikely to originate from the target speech source.

Hearing aids use a number of signal processing techniques to improve performance in noisy environments. These techniques are often grouped into two broad categories:

Directional processing, such as beamforming, exploits spatial differences between microphones to emphasize sounds arriving from certain directions, typically the front.

Digital noise reduction techniques, in contrast, aim to suppress background noise by estimating and attenuating signal components that are unlikely to originate from the target speech source.

Directional processing techniques, like beamforming, have been shown to provide moderate intelligibility benefits for hearing aid users (for speech in the front) of 3-6 dB (Bentler et al. 2004), with binaural implementations sometimes providing a slightly larger improvement in intelligibility at the cost of reduced speech naturalness (Kumar et al. 2023). Digital noise-reduction algorithms, by contrast, have more consistently demonstrated reductions in listening effort and improvements in perceived comfort than improvements in measured speech intelligibility, with intelligibility outcomes varying across studies and test conditions (Lakshmi et al. 2021; Chong and Jenstad 2018).

Recent advances in machine-learning methods have enabled data-driven approaches to speech enhancement that operate on complex acoustic environments. Numerous studies have explored deep neural networks for speech enhancement and separation (Luo and Mesgarani 2019; Isik et al. 2020; Choi et al. 2021; Westhausen and Meyer 2020). However, many of these studies do not measure intelligibility directly, and instead rely on computed metrics like measured SNR, Perceptual Evaluation of Speech Quality (PESQ) (Rix et al. 2001) or Short-time Objective Intelligibility (STOI) (Taal et al. 2010) to evaluate effectiveness. Although these metrics can be informative, improvements in objective measures do not necessarily translate into improved behavioral speech intelligibility for hearing-impaired listeners (Hu and Loizou 2007) and may even reduce intelligibility (Gelderblom 2017). For this reason, direct measurement of speech intelligibility remains essential to evaluate hearing-aid signal-processing strategies.

Of the studies that report intelligibility improvements using machine learning-based methods, a majority have explored longer latencies (20 - 75 ms) than are typical in hearing aids (Healy et al. 2013; Chen et al. 2016; Healy et al. 2015; Healy et al. 2021; Diehl et al. 2023a; Diehl et al. 2023b). Others have a latency suitable for a hearing aid and have been shown to improve intelligibility in offline evaluation (i.e. users listening to audio clips with headphones) (Gaultier and Goehring, 2024; Cornell et al. 2022; Liu and Zhang, 2022; Ouyang et al. 2022; Lei et al. 2022) or in lab evaluations using wired prototypes (Westhausen et al. 2024), but these generally do not run on a wireless hearing aid due to the computational requirements of these models. Two small trials (N=10-11) using a commercial hearing aid running an artificial intelligence (AI)-based noise reduction model have reported an improvement in word recognition in noise with cochlear implant patients and cochlear implant candidates with moderate-to-profound hearing loss (Kolberg et al. 2025, Saoji et al. 2025). The present study reports on the achievement of a large intelligibility improvement over a hearing aid employing AI-based noise reduction. Moreover, the reported improvement was observed across a wide range of hearing loss profiles (including mild hearing loss).

The present study evaluates the intelligibility impact of a spatially guided neural-network algorithm (Spatial AI) implemented in the Fortell AI hearing aids (Fortell Research, New York, NY, USA). The algorithm is designed to preferentially preserve speech arriving from the front while attenuating competing sound sources originating from other directions. This algorithm is meant for situations where the voices of interest are in front of the user, and can be accessed in a hearing aid program called "Front Voices". Unlike conventional beamforming or single-channel noise-reduction approaches, this processing strategy integrates spatial information derived from multiple microphones with data-driven suppression of competing sources. More specifically, the algorithm attenuates (1) competing talkers in the back plane and (2) non-speech noise from all directions.

The objective of this study was to assess speech intelligibility using this spatially guided neural-network processing mode in a controlled multitalker configuration (i.e., "the cocktail party problem") that is often very challenging for beamforming and even machine learning-based approaches. Specifically, target speech was presented from the frontal direction while competing speech originated from the rear hemisphere. Performance was evaluated behaviorally using a standardized speech-in-noise test under multiple challenging signal-to-noise ratios.

This study compares the impact of two different hearing aids– (1) an investigational hearing aid powered by Spatial AI and (2) a commercially available control hearing aid employing neural-network-based noise-reduction processing. A double-blind, randomized, within-subject crossover design was used in which participants completed the same type of speech-in-noise test wearing each of the two devices. The primary, pre-registered endpoint was the odds ratio (OR) for correct word identification between the two hearing aids, for each of three SNR conditions. The study was pre-registered with the Open Science Foundation and the pre-registration can be found at https://doi.org/10.17605/OSF.IO/SN5HW.

Methods

Participants and Setup

30 adult hearing aid wearers with mild to moderately-severe sensorineural hearing loss were recruited for the study. All participants provided written informed consent prior to participation and were compensated for their time. The study was approved by an Institutional Review Board. With N = 30, the crossover design provides ~80% power to detect an OR advantage of roughly 1.6 (Bonferroni-adjusted $\alpha = 0.0167$). An OR = 1.6 corresponds, for example, to an increase from 50% to approximately 60% words correct at a given SNR, an improvement that was considered clinically meaningful.

All participants completed a hearing health history to exclude those with significant otologic conditions, such as infection, pain, or sudden hearing changes. Hearing loss was determined using standard audiometric procedures, including otoscopic examination, a complete audiologic evaluation, and oto-admittance. Assessments were conducted in a quiet room using calibrated equipment, using a Shoebox web-based audiometer. Testing was completed by an audiologist and adhered to established good clinical practice guidelines to ensure accuracy and reliability. Participants were required to have a four-frequency pure-tone average (0.5, 1, 2, and 4 kHz) between 20 dB and 70 dB HL in their better ear. Inclusion criteria also required a word recognition score in quiet of at least 60% and fluency in English. Furthermore, participants completed a QuickSIN (Etymotic Research, Elk Grove Village, IL, USA) test reference (two lists) unaided (these lists were excluded from subsequent testing). Participants were not included or excluded on the basis of their SNR loss.

Table 1 shows the distribution of hearing losses of the study participants. The median pure tone average was 47 dB HL, representing a moderate hearing loss. The median word recognition score was 90%, representing excellent word recognition. The median SNR-loss was 9.5 dB, representing moderate SNR loss. In all cases though, there was a wide range of test scores spanning the range of acceptance criteria for the study. Table 1 shows the minimum and maximum scores among participants, along with the 20th, 50th and 80th percentiles for each of the hearing test results. Figure 1 shows the distribution of audiograms (left ear shown).

Table 1. Participant Hearing Evaluation Data (N=30)

	Min	20%	50%	80%	Max
Age	31	66	80	83	87
Pure Tone Average Air conduction, both ears Avg (500, 1k, 2k, 4k), dB HL	31	42	47	55	65.5
Word Recognition NU-6 % Correct	65%	80%	90%	100%	100%
QuickSIN Score 2 lists dB SNR Loss	1.5	5	9.5	16.5	21

Table 1: The distribution of hearing losses among participants. The minimum, maximum, 20th 50th and 80th percentiles of scores are shown.

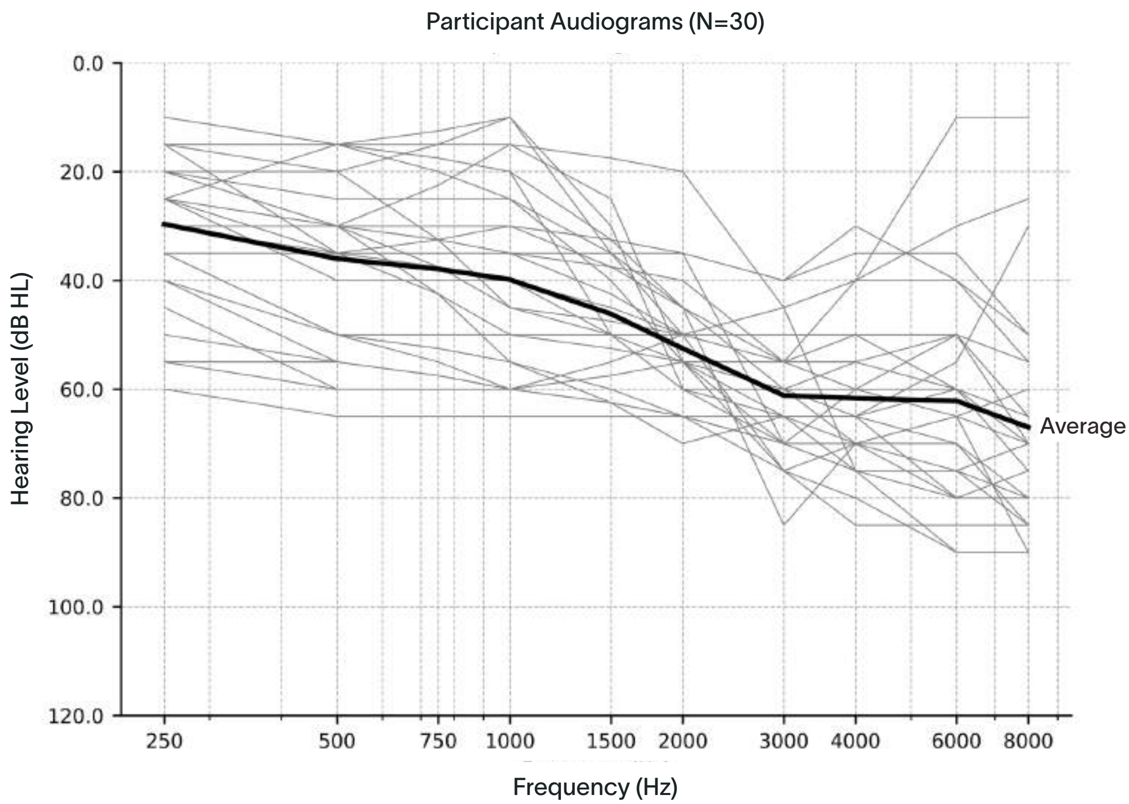


Figure 1. Participants' air-conduction audiograms (left ear shown for each participant), showing range of hearing losses.

Five participants were excluded after their hearing evaluations, two because word recognition was too poor, one because hearing loss was too severe, one because the asymmetry between ears was too large and one due to impacted cerumen. All participants wore different models of hearing aids from the hearing aid models being evaluated in the study. None of the participants wore the investigational nor the control hearing aid.

Hearing Aids

This study compared two sets of hearing aids, an investigational hearing aid (investigational HA) running the Spatial AI model and a control hearing aid (control HA).

The control hearing aid was a receiver-in-canal device from a major manufacturer. It was released in 2024 and represents the highest tech level for its product line. It employed a commercially available neural-network-based noise-reduction algorithm implemented on a dedicated processing chip. For the purposes of this study, the device was programmed using the manufacturer's fitting software such that both directionality and noise-reduction features were set to their highest available levels.

The investigational hearing aid used in the comparison was also a receiver-in-canal hearing aid and implemented a spatially guided neural-network processing mode (Spatial AI) designed to preferentially preserve speech arriving from the front (within an angle of ±45 degrees) while attenuating competing sound sources originating from other directions, including non-speech noise from all directions and interfering speech from the back plane. Processing relied on information derived from multiple microphones integrated into the device.

The neural-network model was trained using supervised, data-driven techniques on a combination of simulated and recorded acoustic mixtures. Simulated training data consisted of mixtures of speech and noise rendered using a proprietary spatial sound simulation pipeline, allowing the model to receive spatialized microphone signals corresponding to different source locations. In this context, clean speech refers to the target talker's signal prior to mixing with competing sources and noise. Additional training data were derived from real-world recordings to expose the model to a range of acoustic conditions, including reverberation, source motion, and head movements. During training, model parameters were optimized (using backpropagation and stochastic gradient descent) to estimate the target speech arriving from the frontal direction. The model was trained over millions of audio samples representing years of audio data. None of the speech materials used for behavioral testing were included in the training data.

The model is significantly larger in terms of memory and computational power than can be run on typical digital signal processing platforms found in hearing aids, so the hearing aid uses a custom designed machine learning coprocessor to run the neural network, which can run up to 100 billion operations per second. The coprocessor is integrated inside the wireless hearing aid and is not part of an external device. This allows the hearing aid to run the neural network continuously without meaningfully compromising battery life or size (when the AI is on, the investigational HA has shorter acoustic latency, longer battery life and is physically smaller than the control HA).

Hearing Aid Fitting

All hearing aids were fitted individually for each participant by a licensed audiologist prior to testing. Devices were programmed using the manufacturers' proprietary fitting software and adjusted to match prescriptive targets based on the NAL-NL2 formula (National Acoustic Laboratories, Sydney, Australia). Real-ear verification was performed using a Verifit 2 system (Audioscan, Dorchester, ON, Canada) to confirm gain targets across frequencies, with the international speech test signal (ISTS) as the stimulus. For both the investigational and control hearing aids, gain deviations were adjusted to fall within ±5 dB of prescriptive targets across the 500–4000 Hz frequency range. This procedure was repeated for each participant to ensure comparable audibility between devices prior to speech-in-noise testing.

For both hearing aid types, power domes were used to provide a consistent acoustic coupling and to minimize the contribution of the unamplified direct sound path during testing. No participants reported discomfort or occlusion-related complaints during the study. Directionality and noise-reduction features were enabled for both devices and configured to their highest available settings, consistent with manufacturer-recommended configurations for challenging listening environments, in line with the study objective of evaluating performance under difficult conditions. Fitting procedures and verification steps were identical for both hearing aid types and were completed prior to randomization and blinding.

Speech-in-Noise Testing

Speech-in-noise testing constituted the primary endpoint of the study. Testing was conducted across three sessions on separate days in a recording studio in New York City. Data collection procedures, including maintenance of double blinding and scoring accuracy, were overseen by an academic research collaborator independent of device development, who was present on site for the majority of testing sessions.

Speech intelligibility was assessed using the Quick Speech-in-Noise (QuickSIN) test, a standardized and widely used clinical measure of sentence recognition in multitalker babble. The QuickSIN provides a controlled method for estimating speech intelligibility across a range of signal-to-noise ratios (SNRs) and has been validated for use in listeners with hearing loss. Its use here allowed for direct comparison of performance across hearing aid conditions under identical acoustic and procedural constraints. Among the various, validated speech-in-noise tests to choose from, the QuickSIN test was particularly attractive because it represents a challenging environment where the distracting noise is other people talking. Note that even though the Spatial AI model was trained to remove non-speech noise (in addition to competing speakers), the present study only focuses on the case of multi-talker noise and does not include non-speech noise.

Test materials were unaltered from the standard QuickSIN and consisted of a target speaker producing complete sentences and a competing track containing four simultaneous talkers ("babble"). To create the test, 10 QuickSIN lists (representing 60 sentences) were split into 4 blocks of 3, 3, 2 and 2 lists respectively (henceforth referred to as Blocks 1, 2, 3 and 4). Within each block, participants heard an equal number of sentences for each SNR. Each sentence was presented at one of three SNRs: 6 dB, 0 dB, and -6 dB. The speech presentation level was 65 dB SPL and the noise presentation level was changed to achieve the target SNR (calibrated using a reference microphone behind the ear).

Testing order was counterbalanced across participants as part of the crossover design. Participants were randomly assigned to one of 2 groups (A and B). Group A started with the investigational HA in Block 1, then switched to control HA for Block 2, then back to investigational HA for Block 3, then back to control HA for Block 4. Group B completed the Blocks in the same order, but wore the control HA for Blocks 1 and 3 and the investigational HA for Blocks 2 and 4. Table 2 specifies the SNR and number of sentences for each Block and which hearing aid was worn for each Block.

Table 2. Sentence Blocks by SNR and Treatment Assignment

Block	SNR	Sentences	Group A Treatment	Group B Treatment
Block 1	6 dB	6	Investigational HA	Control HA
	0 dB	6	Investigational HA	Control HA
	-6 dB	6	Investigational HA	Control HA
Block 2	6 dB	6	Control HA	Investigational HA
	0 dB	6	Control HA	Investigational HA
	-6 dB	6	Control HA	Investigational HA
Block 3	6 dB	4	Investigational HA	Control HA
	0 dB	4	Investigational HA	Control HA
	-6 dB	4	Investigational HA	Control HA

Block 4	6 dB	4	Control HA	Investigational HA
	0 dB	4	Control HA	Investigational HA
	-6 dB	4	Control HA	Investigational HA

Table 2: The order of the sentence blocks for groups A and B.

Testing was conducted in a sound-treated room to ensure precise control over stimulus presentation and spatial configuration. Participants were seated in the middle of a 4-loudspeaker array (Genelec 8010s, Genelec Oy, Iisalmi, Finland), as shown in Figure 2, with loudspeakers positioned at 0 azimuth (the "front loudspeaker"), 135 azimuth, 180 azimuth and -135 azimuth (the 3 "back loudspeakers"). Loudspeakers were each positioned 1.2 meters from the head of the participant at eye level. This configuration was selected to provide a controlled and repeatable multitalker scenario that isolates the effects of frontal-target, rear-interferer geometry, rather than to simulate the full spatial and acoustic complexity of everyday listening environments.

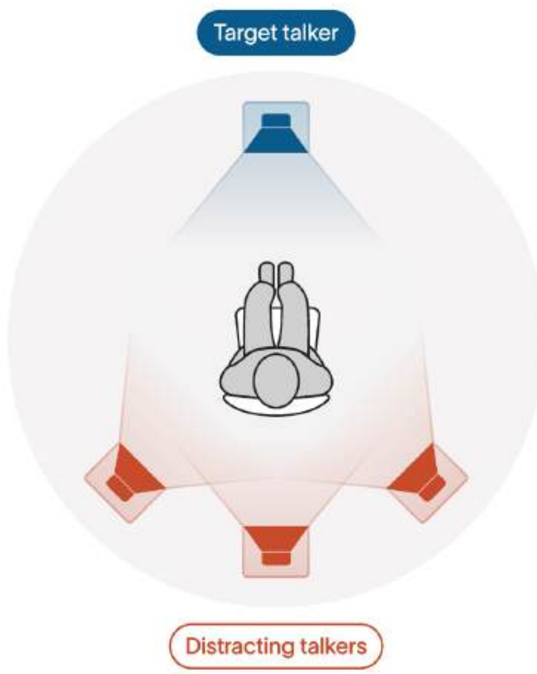


Figure 2: Diagram illustrating the spatial configuration of speech and noise sources. Target speech was presented from the front loudspeaker (0° azimuth), while multitalker babble was presented from one of three loudspeakers behind the participant. The specific rear loudspeaker varied pseudorandomly across sentences following the same pattern for all participants

Neither the loudspeakers nor the chair were moved during the testing. The front loudspeaker played the target speech while the noise track was played out of one of the back loudspeakers. Which loudspeaker the noise played out of varied randomly with every sentence, though the test followed the same pattern of loudspeaker selection for every participant. After a researcher conveyed instructions, participants heard 3 practice sentences (one at each of the test SNRs) to confirm that they understood the instructions. Participants then completed the Speech-in-Noise test in one sitting.

Randomization and Blinding

Participants were randomly assigned to Group A or Group B (balanced so that both Group A and Group B would have 15 participants each). The scorer was seated behind a screen so that they could hear the participants' responses but not see the participants or know which of the two hearing aids they were wearing. An academic research collaborator also scored responses for 5 of the 30 participants to confirm scoring consistency.

Participants were told that the purpose of the study was to compare two types of hearing aid processing, with no other details about the manufacturer or processing provided. To maintain blinding for both the participant and the research team, the audiologist was the only one handling the hearing aids. For each of the hearing aid switches, the audiologist came into the testing room and stood behind the participant to take one pair of hearing aids off and put the other pair on. The audiologist was not blinded but they had no other role in test administration. At no point were participants informed which device was on their ear or able to see the device being used. The audiologist also ensured that the physical fit was secure and occlusive for both hearing aids. Participants did not handle the devices at any point during testing, and all device changes were performed behind the participant to minimize visual or tactile cues.

Lab-measured Objective Metrics

Objective acoustic metrics were obtained to complement the behavioral speech intelligibility results measured in human participants. Specifically, output signal-to-noise ratios (SNRs) for the investigational and control hearing aids were estimated using the Hagerman–Olofsson phase-inversion method (Hagerman et al., 2004) applied to the QuickSIN stimuli. This method allows separation of speech and noise components from simultaneously presented signals and enables quantitative estimation of output SNR under controlled conditions.

Let s and n denote the speech and noise components of the stimuli presented through the loudspeaker array. The Hagerman-Olofsson method involves making three recordings of the output of the hearing aid:

$$m_o = s + n$$
$$m_i = s - n$$
$$m_\epsilon = -s - n$$

The estimated speech and noise components of the hearing aid output, S and N, can then be reconstructed as:

$$S = \left(m_o + m_i\right) / 2$$
$$N = \left(m_o - m_i\right) / 2$$

These components can then be used to compute the SNR of the hearing aid output:

$$SNR = 10\, log_{10}\left(< S^2 > \;/\; < N^2 >\right).$$

The quantity

$$E = \left(m_{\epsilon} + m_{o}\right) / 2$$

is used to estimate the error in the speech and noise reconstruction. The Hagerman error of the speech and noise estimate, respectively, can then be computed as:

$$\epsilon_{s} = 10\, log_{10}\left(< E^{2} > \; / \; < S^{2} >\right)$$
$$\epsilon_{n} = 10\, log_{10}\left(< E^{2} > \; / \; < N^{2} >\right).$$

When computing these quantities, speech intelligibility index (SII) weighting was applied.

To compute the Hagerman-Olofsson SNR and HASPI v2 objective metrics, the QuickSIN test was recorded on a KEMAR acoustic manikin (GRAS Sound & Vibration, Holte, Denmark) fitted with the investigational and control HAs. The test was recorded in precisely the same configuration as the test with the human subjects. The KEMAR was placed in the middle of the speaker array (each speaker was 1.2 meters from the KEMAR at eye level). The QuickSIN test was also unchanged from the test with the human subjects: it used the same three input SNRs (-6 dB, 0 dB and 6 dB). The speech presentation level was 65 dB SPL and the noise was adjusted to achieve the desired SNR. Both HAs were programmed to NAL-NL2 targets for the N3 audiogram and power domes were used to provide a consistent acoustic seal. Testing was performed in a standard office room (approximately 20 square meters in area) with a reverberation time (RT-60) of 450 ms. As in the intelligibility test with human subjects, both investigational and control HAs were configured to maximize both directionality and AI-based noise reduction. To enable Hagerman–Olofsson measurements, feedback cancellation, impulse noise reduction, and wind-noise reduction features were disabled on both devices.

Results

Speech Intelligibility Testing

Speech intelligibility scores were computed as the percentage of target words correctly identified, with five target words per sentence. Figure 3 shows the percent correct for each hearing aid condition at each SNR, with 95% confidence intervals. At the most challenging SNR of -6 dB, mean word recognition increased from 7% with the control HA to 50% with the investigational HA. At SNRs of 0 dB and 6 dB, the percent correct increased from 31% to 80% and 69% to 90%, respectively.

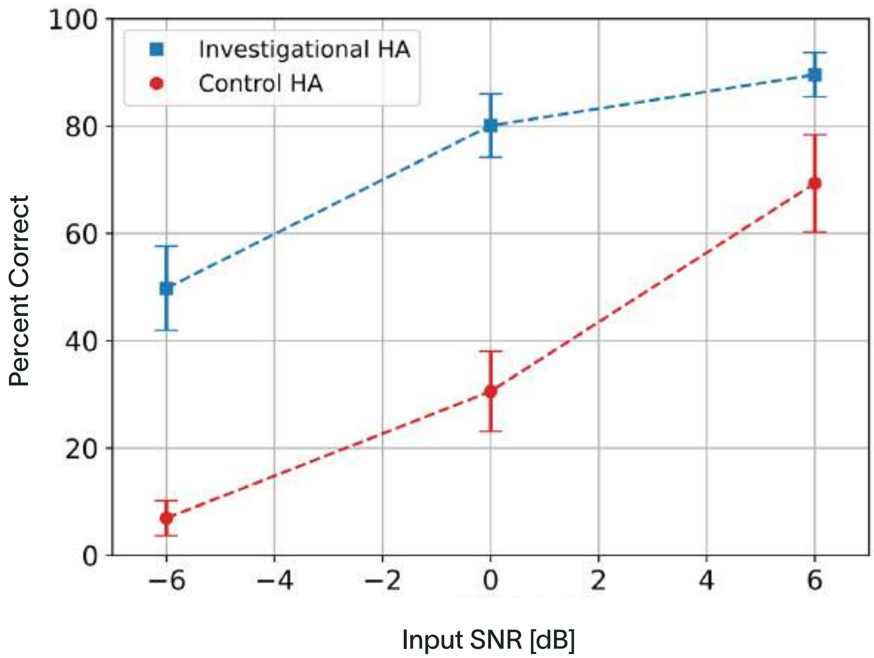


Figure 3: Average percent of words identified correctly for each hearing aid type for each SNR. Error bars show 95% confidence intervals.

We fitted the data using a generalized linear mixed-effects model with a logistic link function. Hearing aid condition, SNR, and testing order were treated as fixed effects, while the participant was treated as a random effect. An interaction between the hearing aid and SNR was included to test if the benefit of one hearing aid over the other varied with SNR. Statistical significance of the hearing-aid effect was evaluated using Wald tests on fixed-effect coefficients and a two-tailed p-value was computed. To account for multiple comparisons across the three SNR conditions, a Bonferroni correction was applied (significance threshold of $p<0.0167$) for comparisons at individual SNR levels.

As shown in Table 3, the investigational HA increased the odds of correct word identification relative to the control hearing aid at all tested SNRs. The odds ratio was 18.9 at –6 dB, 10.6 at 0 dB, and 6.0 at +6 dB SNR. All effects were highly statistically significant.

Table 3. Investigational HA versus Control HA Percent Correct Odds Ratios for each SNR

SNR	Investigational vs Control Odds Ratio	95% Confidence Interval	*p* (two-tailed)
-6 dB	18.9	[16.7, 21.3]	p<<0.0001
0 dB	10.6	[9.8, 11.5]	p<<0.0001
+6 dB	6.0	[5.3, 6.8]	p<<0.0001

Table 3: The odds ratio of getting words correct for the investigational HA vs. the control HA for each of the three SNRs.

The fitted model was used to estimate the SNR corresponding to 50% word recognition (SNR-50) for each hearing aid condition by solving the logistic function for a predicted probability of 0.5. Using this approach, the estimated SNR-50 for the control hearing aid was 3.08 dB (95% CI [2.84, 3.31]), whereas the estimated SNR-50 for the investigational hearing aid was –6.15 dB (95% CI [–6.86, –5.45]), corresponding to an estimated SNR-50 improvement of 9.2 dB. These estimates provide a compact summary of the shift in the psychometric function observed across the tested SNR range. Because estimates were derived from performance measured at three discrete SNRs, SNR-50 values should be interpreted as model-based summaries rather than precise thresholds.

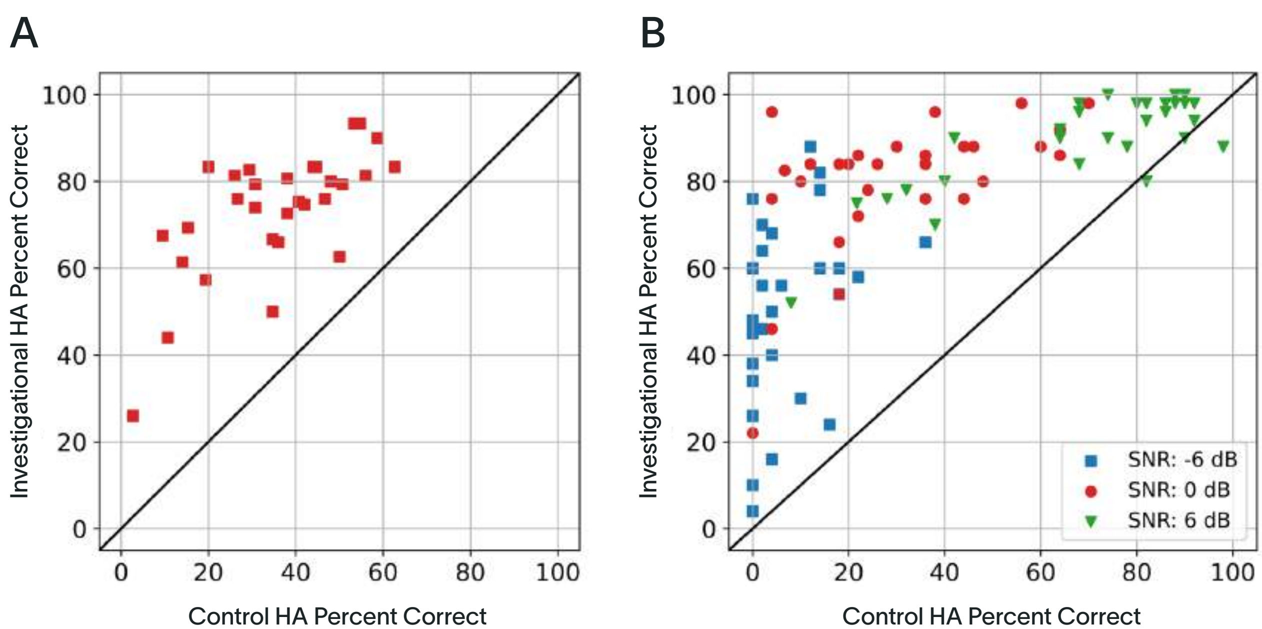


Figure 4: Panel A shows the percentage of words identified correctly wearing the investigational hearing aid and the percentage of words identified correctly wearing the control hearing aid for each participant averaged across all tested SNRs (n=30). Figure 4b shows the same data broken down by SNR (n=30 for each SNR).

Figures 4a and 4b show results for each participant. Figure 4a shows percent-correct scores averaged across SNRs for each participant, comparing investigational and control hearing aid conditions. All participants demonstrated higher average performance with the investigational hearing aid, the improvement being statistically significant for every individual. Figure 4b shows the same data broken down by SNR. At the most challenging SNR (SNR = -6 dB), one third of participants got 0 words correct using the control HA and no participant exceeded 50% correct, whereas with the investigational hearing aid no participant scored 0% and half of participants had scores above 50% correct.

The test protocol was inadvertently misapplied for one participant, such that they completed 36 sentences (Blocks 1 and 2) with the control hearing aid followed by 24 sentences (Blocks 3 and 4) with the investigational hearing aid. Because the distribution of SNRs of the four Blocks was identical, performance should be unaffected and the participant was included in the study results. Excluding this participant yielded only minor changes: the odds ratios were 18.2, 10.2 and 5.8 at SNRs of -6 dB, 0 dB and 6 dB, respectively, and SNR-50 improvement was 9.0 dB.

All results for the study are available at https://doi.org/10.17605/OSF.IO/9EFV3.

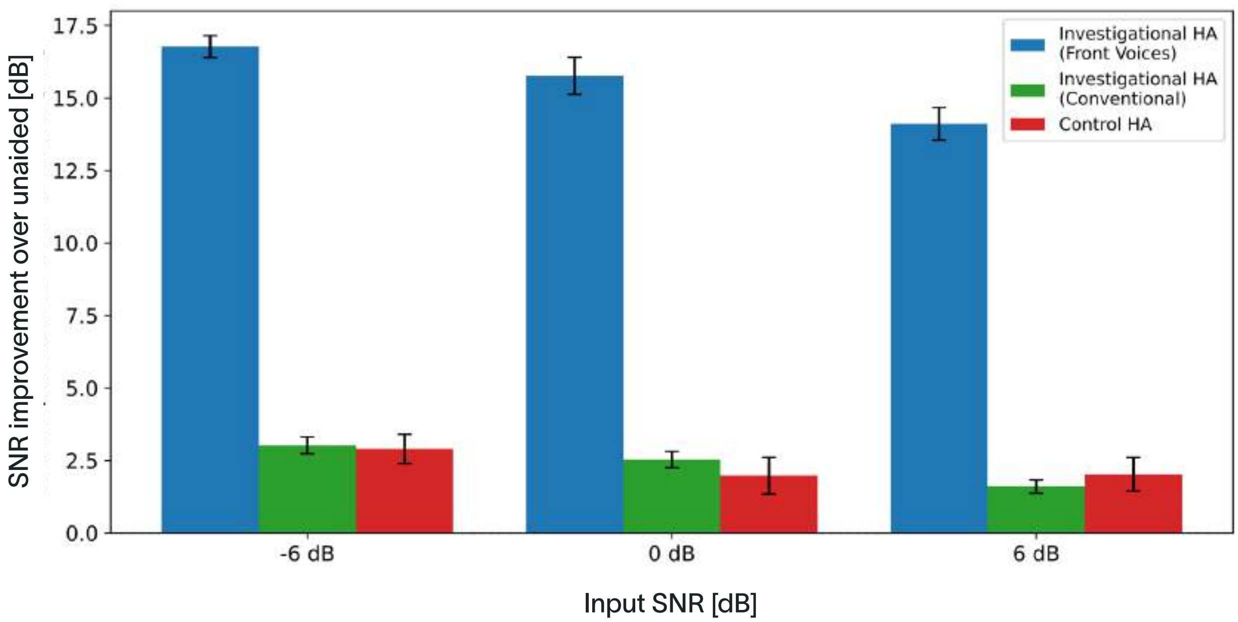


Figure 5: The SII-weighted SNR improvement (over the unaided condition) at the three input SNRs for the investigational HA (Front Voices mode, which uses Spatial AI), the control HA and Conventional program (no AI) on the investigational HA. Error bars show 95% confidence intervals.

Lab-measured Objective Metrics

Figure 5 shows SII-weighted output SNR improvements relative to the unaided condition for the investigational hearing aid, the control hearing aid, and a conventional (non-AI) processing program on the investigational device. This conventional program includes the standard suite of hearing aid algorithms, including classical noise reduction, adaptive beamforming and wide dynamic range compression (WDRC). Averaged across input SNRs, the investigational hearing aid operating in the spatially guided mode achieved an SII-weighted SNR improvement of 15.5 dB, compared with 2.3 dB for the control hearing aid. The conventional program achieved an SNR improvement of 2.4 dB, similar to that of the control device.

In terms of unweighted SNR, the improvements over unaided were 17.1 dB for the investigational hearing aid, 1.9 dB for the control hearing aid, and 2.6 dB for the conventional program. The Hagerman speech error was -22.2 dB, -13.7 dB and -16.8 dB for the investigational HA (Front Voices), control HAs and the investigational HA (Conventional), respectively. The Hagerman noise error was -14.4 dB, -19.1 dB and -22.1 dB for the investigational HA (Front Voices), control HAs and the investigational HA (Conventional), respectively.

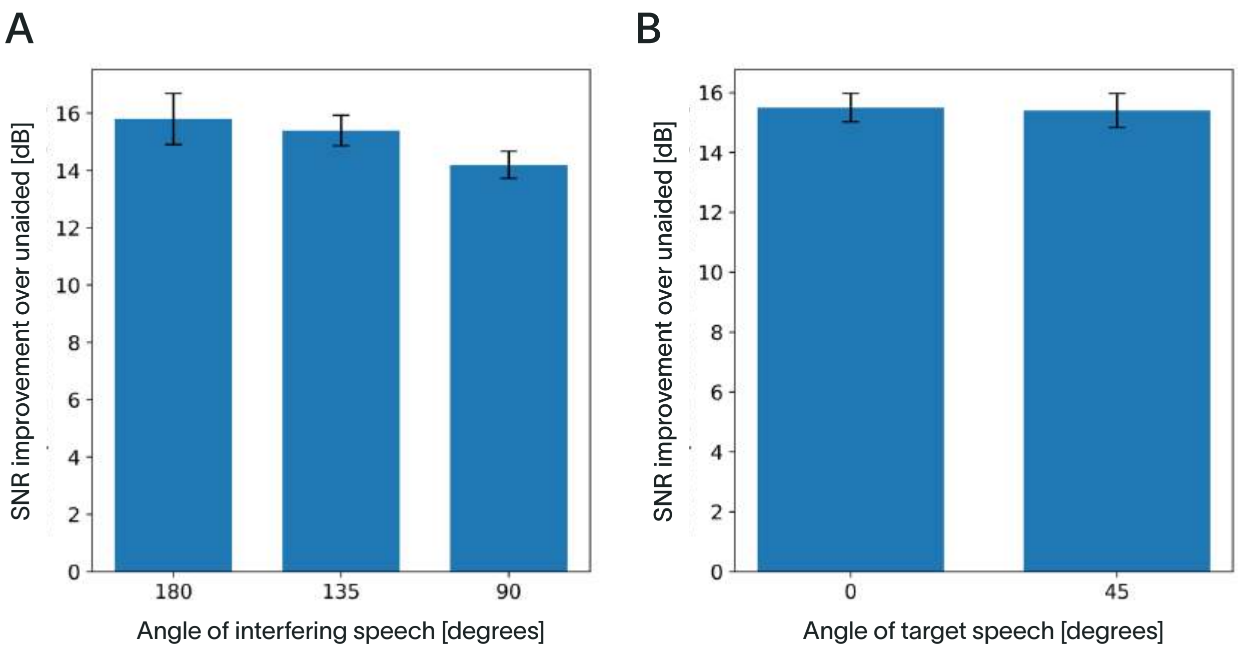


Figure 6: Panel A shows the SII-weighted SNR improvement (over the unaided condition) as the angle of the loudspeaker that plays out the interfering speech is varied (with target speech coming from the front). Panel B shows the SII-weighted SNR improvement (over the unaided condition) as the angle of the loudspeaker that plays out the target speech is varied (with interfering speech in the standard configuration of ±135, 180 degrees). Error bars show 95% confidence intervals.

Figure 6a shows SII-weighted SNR improvement as a function of the azimuth of the interfering speech source, with target speech presented from the front. Averaged across input SNRs, SNR improvement with interfering speech at 90° was 14.2 dB, only 1.6 dB lower than when interfering speech originated from directly behind. Figure 6b shows that SNR improvement remained high when target speech was presented at 45° relative to the front (with interfering speech in the standard configuration of ±135, 180 degrees), with performance similar to that observed for frontal target presentation.

We also computed the Hearing-Aid Speech Perception Index (HASPI) v2 (Kates et al. 2021), as a complementary objective metric. The investigational HA scored 0.24, 0.57 and 0.77 at input SNRs of -6 dB, 0 dB and 6 dB, respectively. Meanwhile the control HA scored 0.02, 0.03 and 0.16. Fitting these scores with logistic curves gives an SNR-50 difference of 15.4 dB between the investigational and control HAs. Again, these HASPI results are reported for descriptive purposes and are not intended to substitute for behavioral intelligibility outcomes.

Discussion

Our results demonstrate substantial speech intelligibility gains in multitalker noise for individuals with mild to moderately-severe hearing loss when using an investigational hearing aid implementing a spatially guided neural-network processing mode, compared with a commercially available control hearing aid employing neural-network-based noise reduction. Across all tested SNRs, behavioral intelligibility was consistently higher with the investigational device, with particularly large benefits observed under the most challenging multitalker conditions. These effects were evident at both the group and individual participant levels.

The magnitude and consistency of the observed effects are notable given the conservative study design. All participants were experienced hearing-aid users, both devices were fitted to prescriptive targets with matched audibility, and testing employed a randomized, double-blind, within-subject crossover design. Importantly, every participant demonstrated improved performance with the investigational device, reducing the likelihood that the observed benefit reflects individual outliers or learning effects.

The observed intelligibility benefits must be interpreted in the context of the specific spatial configuration tested. Target speech was presented from the frontal direction while competing multitalker babble originated from the rear hemifield, a geometry that emphasizes spatial separation between desired and interfering signals. This configuration was selected deliberately to provide a controlled and repeatable multitalker scenario that isolates the impact of spatially selective processing, rather than to simulate the full complexity of everyday listening environments. Accordingly, the present findings do not imply that similar benefits will necessarily be obtained in all real-world listening situations. In particular, performance may differ in environments where competing talkers originate from the frontal hemifield, where sources move dynamically, or where listeners orient away from the target talker. Future work will be needed to characterize performance under such conditions.

Previous studies of directional processing and noise-reduction algorithms have reported modest and variable intelligibility benefits, particularly in multitalker environments. The present results extend this literature by demonstrating that, under controlled spatial conditions, spatially guided neural-network processing can yield substantially larger intelligibility improvements than those typically reported for conventional approaches. Importantly, these gains were demonstrated using real-time, wearable hearing aids and behavioral outcome measures.

From a clinical perspective, the magnitude of the intelligibility improvement observed in this controlled multitalker configuration is notable. In particular, the estimated 9.2 dB improvement in SNR-50 represents a substantial shift in the speech-in-noise operating point relative to the control device. Previous work has shown that considerably smaller SNR improvements can be perceptually meaningful to hearing aid users (McShefferty et al. 2015). Although the present study does not address listener behavior, device adoption, or long-term outcomes, the size of the observed SNR-50 shift suggests that spatially guided processing has the potential to support materially improved speech understanding when favorable spatial separation between target and competing talkers is present.

While the Hagerman-Olofsson SNR and HASPI v2 measurements broadly agreed with each other in terms of capturing the improvement of the investigational HA over the control HA, both metrics predicted an improvement that was larger than what was observed in the speech intelligibility test with human subjects. This further underscores the importance of measuring intelligibility directly, rather than relying solely on objective metrics.

The convergence of behavioral and objective results suggests that the investigational processing mode achieves greater spatial suppression of competing speech under the tested geometry. One plausible contributing factor is the explicit use of spatial information derived from multiple microphones, which provides a reliable cue for separating target speech from competing sources. Under multitalker conditions, such processing may reduce the impact of informational masking, which is known to disproportionately limit human speech perception compared with steady or modulation-dominated noise (Stone et al. 2012). However, the present data do not permit isolation of specific algorithmic mechanisms, and further work will be required to clarify how spatial cues, masking type, listener orientation, and head movement interact in more complex acoustic scenes.

More broadly, these findings underscore the importance of implementation considerations in modern hearing aid signal processing. The ability to deploy machine-learning models that integrate spatial information in real time increasingly depends on dedicated on-device processing resources capable of meeting the power, latency, and form-factor constraints of wearable devices. As hardware capabilities continue to evolve, such architectures may enable more sophisticated combinations of spatial and data-driven processing. Systematic evaluation across a broader range of acoustic environments will be essential to determine how these advances translate into everyday listening benefit.

Limitations and Future Research

This study focused exclusively on speech intelligibility in multitalker noise under a specific and controlled spatial configuration. Competing speech sources were restricted to the rear hemifield and testing was conducted in a low-reverberation environment, conditions that emphasize spatial separation between target and interferers. Performance may differ in environments with greater reverberation, with competing talkers originating from the frontal hemifield, or in dynamic listening situations involving head movements and changing source locations.

The investigational processing mode evaluated here is intended for situations in which target speech is located in front of the listener. Use of spatially selective processing when the target is located laterally, or when competing speech falls within a narrow frontal angle, may reduce rather than improve intelligibility. Characterizing the conditions under which spatially guided processing is beneficial, neutral, or detrimental remains an important area for future investigation.

The present study employed power domes to minimize contributions from the direct acoustic path and to isolate effects attributable to hearing aid processing. Many hearing aid users may prefer more open fittings for everyday wear, and future work should examine how the observed benefits vary with coupling and venting. In addition, intelligibility was assessed at three challenging SNRs; although performance with the investigational device approached 50% even at the lowest SNR, testing at more adverse SNRs may further clarify performance limits. Lastly, while the control device was chosen because it represented state of the art at the time of writing, it would be instructive to compare to other hearing aids using other processing techniques.

Future research should look at the impact of Spatial AI processing on overall wearer satisfaction. People don't live their lives in a noisy restaurant. The multi-talker noise scenario is only representative of a portion of a wearer's day. Therefore it would be interesting to see how a hearing aid with this technology compared to today's top of the line hearing aids when worn throughout the day.

Conclusions

Multi-talker environments pose a particular challenge for hearing aid users, in part because competing speech is both acoustically similar to the target and highly distracting. Under a well-defined multitalker spatial configuration, the present study demonstrates that a spatially guided neural-network hearing aid algorithm can yield large and consistent improvements in speech intelligibility relative to a contemporary control device. Notably, in the most challenging multi-talker environment participants had 18.9x higher odds of understanding speech compared to the top AI hearing aids on the market today. These findings were observed despite a conservative study design involving experienced hearing-aid users, prescriptive fitting with matched audibility, and a randomized, double-blind, within-subject crossover protocol. Together, these results motivate further investigation of spatially selective neural processing across a wider range of acoustic environments and listening behaviors.

References

Bentler R, Wu YH, Kettel J, Hurtig R. Digital noise reduction: outcomes from laboratory and field studies. Int J Audiol. 2008 Aug;47(8):447-60. doi:10.1080/14992020802033091.

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 May;139(5):2604-2612. doi:10.1121/1.4948445.

Choi HS, Park S, Lee JH, Heo H, Jeon D, Lee K. Real-time denoising and dereverberation with tiny recurrent U-Net. In: 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2021. p. 5789–5793. doi:10.1109/ICASSP39728.2021.9414852.

Chong FY, Jenstad LM. A critical review of hearing-aid single-microphone noise-reduction studies in adults and children. Disabil Rehabil Assist Technol. 2018 Aug;13(6):600-608.

Cornell S, Wang ZQ, Masuyama Y, Watanabe S, Pariente M, Ono N. Multi-channel target speaker extraction with refinement: the WAVLab submission to the second Clarity Enhancement Challenge. In: Proceedings of the Clarity-2022 CEC2 Workshop. Clarity Project; 2022. Available from: https://claritychallenge.org/clarity2022-CEC2-workshop/papers/Clarity_2022_CEC2_paper_cornell.pdf.

Diehl PU, Singer Y, Zilly H, Schönfeld U, Meyer-Rachner P, Berry M, Sprekeler H, Sprengel E, Pudszuhn A, Hofmann VM. Restoring speech intelligibility for hearing aid users with deep learning. Sci Rep. 2023 Feb 15;13(1):2719. doi:10.1038/s41598-023-29871-8.

Diehl PU, Zilly H, Sattler F, Singer Y, Kepp K, Berry M, Hasemann H, Zippel M, Kaya M, Meyer-Rachner P, Pudszuhn A, Hofmann VM, Vormann M, Sprengel E. Deep learning-based denoising streamed from mobile phones improves speech-in-noise understanding for hearing aid users. Frontiers in Medical Engineering. 2023;1:1281904. doi:10.3389/fmede.2023.1281904.

Gaultier C, Goehring T. Recovering speech intelligibility with deep learning and multiple microphones in noisy-reverberant situations for people using cochlear implants. J Acoust Soc Am. 2024 Jun;155(6):3833-3847. doi:10.1121/10.0026218.

Gelderblom FB, Tronstad TV, Viggen EM. Subjective intelligibility of deep neural network-based speech enhancement. In: Proc Interspeech 2017. 2017;1968-1972. doi:10.21437/Interspeech.2017-1041.

Hagerman B, Olofsson A. A method to measure the effect of noise reduction algorithms using simultaneous speech and noise. Acta Acust United Acust. 2004;90:356-361.

Healy EW, Tan K, Johnson EM, Wang D. An effectively causal deep learning algorithm to increase intelligibility in untrained noises for hearing-impaired listeners. J Acoust Soc Am. 2021;149:3943-3953.

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 Sep;138(3):1660-1669. doi:10.1121/1.4929493.

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 Oct;134(4):3029-3038. doi:10.1121/1.4820893.

Hu Y, Loizou PC. A comparative intelligibility study of speech enhancement algorithms. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2007;4:561–564.

Isik U, Giri R, Phansalkar N, Valin JM, Helwani K, Krishnaswamy A. PoCoNet: better speech enhancement with frequency-positional embeddings, semi-supervised conversational data, and biased loss. In: Proc Interspeech 2020. 2020. p. 2487–2491. doi:10.21437/Interspeech.2020-3027.

Kates JM, Arehart KH. The Hearing-Aid Speech Perception Index (HASPI) version 2. Speech Commun. 2021;131:35–46. doi:10.1016/j.specom.2020.05.001.

Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, Banerjee S. Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. J Acoust Soc Am. 2004 Oct;116(4 Pt 1):2395-2405. doi:10.1121/1.1784440. Erratum in: J Acoust Soc Am. 2006 Mar;119(3):1888.

Kolberg C, Holbert SO, Bogle JM, Saoji AA. DNN-based noise reduction significantly improves bimodal benefit in background noise for cochlear implant users. J Clin Med. 2025;14(15):5302. doi:10.3390/jcm14155302.

Kumar S, Guruvayurappan A, Pitchaimuthu AN, Nayak S. Efficacy and effectiveness of wireless binaural beamforming technology of hearing aids in improving speech perception in noise: a systematic review. Ear Hear. 2023 Nov-Dec;44(6):1289-1300. doi:10.1097/AUD.0000000000001374.

Lakshmi MSK, Rout A, O'Donoghue CR. A systematic review and meta-analysis of digital noise reduction hearing aids in adults. Disabil Rehabil Assist Technol. 2021 Feb;16(2):120-129. doi:10.1080/17483107.2019.1642394. Epub 2019 Sep 10.

Lei T, Hou Z, Hu Y, Yang W, Sun T, Rong X, Wang D, Lu J. A hybrid beamforming and post-processing system for the second Clarity Enhancement Challenge. In: Proceedings of the Clarity-2022 CEC2 Workshop. Clarity Project; 2022. Available from: https://claritychallenge.org/clarity2022-CEC2-workshop/papers/Clarity_2022_CEC2_paper_lei.pdf.

Liu J, Zhang X. DRC-NET for the 2nd Clarity Enhancement Challenge. In: Proceedings of the Clarity-2022 CEC2 Workshop. Clarity Project; 2022. Available from: https://claritychallenge.org/clarity2022-CEC2-workshop/papers/Clarity_2022_CEC2_paper_liu.pdf.

Luo Y, Mesgarani N. Conv-TasNet: surpassing ideal time-frequency magnitude masking for speech separation. IEEE/ACM Trans Audio Speech Lang Process. 2019 Aug;27(8):1256-1266. doi:10.1109/TASLP.2019.2915167.

McShefferty D, Whitmer WM, Akeroyd MA. The just-noticeable difference in speech-to-noise ratio. Trends Hear. 2015 Feb 12;19:2331216515572316. doi:10.1177/2331216515572316.

Ouyang C, Fei K, Zhou H, Li L, Lu C. A low-latency deep learning system for the second Clarity Enhancement Challenge. In: Proceedings of the Clarity-2022 CEC2 Workshop. Clarity Project; 2022. Available from: https://claritychallenge.org/clarity2022-CEC2-workshop/papers/Clarity_2022_CEC2_paper_ouyang.pdf.

Rix AW, Beerends JG, Hollier MP, Hekstra AP. Perceptual evaluation of speech quality (PESQ) – a new method for speech quality assessment of telephone networks and codecs. Proc IEEE Int Conf Acoust Speech Signal Process (ICASSP). 2001;2:749–752.

Saoji AA, Sheikh BA, Bertsch NJ, Goulson KR, Graham MK, McDonald EA, Bross AE, Vaisberg JM, Kuhnel V, Voss SC, Qian J, Hogan CH, DeJong MD. How does deep neural network-based noise reduction in hearing aids impact cochlear implant candidacy? Audiol Res. 2024 Dec 13;14(6):1114–1125. doi:10.3390/audiolres14060092.

Soli SD, Wong LLN. Assessment of speech intelligibility in noise with the Hearing in Noise Test. Int J Audiol. 2008;47(6):356-361. doi:10.1080/14992020801895136.

Stone MA, Füllgrabe C, Moore BCJ. Notionally steady background noise acts primarily as a modulation masker of speech. J Acoust Soc Am. 2012;132:317-326.

Taal CH, Hendriks RC, Heusdens R, Jensen J. A short-time objective intelligibility measure for time-frequency weighted noisy speech. Proc IEEE Int Conf Acoust Speech Signal Process (ICASSP). 2010;pp. 4214–4217. doi:10.1109/ICASSP.2010.5495701.

Vermiglio AJ, Herring CC, Heeke P, Post CE, Fang X. Sentence recognition in steady-state speech-shaped noise versus four-talker babble. J Am Acad Audiol. 2019 Jan;30(1):54-65. doi:10.3766/jaaa.17083.

Westhausen NL, Meyer BT. Dual-signal transformation LSTM network for real-time noise suppression. In: Proc Interspeech 2020. 2020. p. 2477–2481. doi:10.21437/Interspeech.2020-2631.

Westhausen NL, Meyer BT, Doclo S. Real-time multichannel deep speech enhancement in hearing aids: comparing monaural and binaural processing in complex acoustic scenarios. J Acoust Soc Am. 2024;155(4):2326-2340. doi:10.1121/10.0026123.

Wilson RH, Carnell CS, Cleghorn AL. The Words-in-Noise (WIN) test with multitalker babble and speech-spectrum noise maskers. J Am Acad Audiol. 2007 Jun;18(6):522-9. doi:10.3766/jaaa.18.6.7.

fortell