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 direction while suppressing 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 control hearing aid, which also ran a real-time neural network noise reduction algorithm and represented today's state of the art. Both devices were set to maximize directionality and noise reduction. 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, suggesting broad applicability across hearing loss profiles. 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 suggest Spatial AI will play a meaningful role in addressing the cocktail party problem for hearing-impaired listeners, with effect sizes greatly exceeding

established thresholds for clinical significance. The results indicate that spatially-aware neural network processing may offer a new technological paradigm for hearing assistance in complex acoustic environments.

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 noise for hearing aid wearers. Typically these techniques fall into two categories:

1. Directional noise reduction, whereby algorithms like beamforming selectively pick up sounds coming from the front of the wearer and
2. Digital noise reduction techniques, which typically estimate the noise level (or SNR) across frequency bands and use this information to control a time-varying gain function that selectively reduces background noise while preserving speech.

Directional techniques like beamforming have been shown to provide moderate intelligibility benefits for hearing aid wearers (for speech in the front) of 3-6 dB (Bentler et al. 2004), with binaural beamforming providing a slightly larger improvement in intelligibility at the cost of less speech naturalness, as shown in meta reviews (Kumar et al. 2023). On the other hand, meta-analyses (Lakshmi et al. 2021; Chong and Jenstad 2018) have found that digital noise reduction does not consistently offer significant speech intelligibility benefits, though some individual studies have reported meaningful effect sizes.

Numerous studies have explored deep neural networks for speech enhancement (Luo and Mesgarani 2019; Isik et al. 2020; Choi et al. 2021; Westhausen and Meyer 2020), yet 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. It is widely known that many algorithms improve objective metrics and yet do not improve intelligibility (Hu and Loizou 2007) or even reduce intelligibility (Gelderblom 2017). For this reason it’s critical to measure speech intelligibility directly rather than assume that measured SNR improvements will translate into real-world intelligibility benefits.

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).

This study examines the intelligibility impact of the Fortell AI hearing aids (Fortell Research, New York, NY, USA) that run a spatial neural network-based algorithm (Spatial AI) that is trained to isolate voices in front of a hearing aid wearer and attenuate both (1) competing talkers in the back plane and (2) non-speech noise from all 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", which users can select via a mobile app or via a physical button on the device. Spatial AI is neither beamforming nor single-channel noise reduction, and so does not fit cleanly into the categorization that is typically used for these techniques. The algorithm takes as input information derived from multiple internal microphones and outputs an enhanced version of the audio that isolates speech from in front of the user.

The Spatial AI model has the potential to improve listening in the most challenging listening situations, including in multi-talker noise (ie "the cocktail party problem"). In these environments, beamforming is often not sufficient to alleviate listening difficulties and existing noise reduction techniques, even machine learning based approaches, struggle to remove noise because they cannot differentiate between wanted and unwanted speech. The Spatial AI model evaluated here uses spatial information to filter unwanted speech based on its position relative to the wearer.

This study compares the impact of two different hearing aids– (1) an investigational hearing aid powered by Spatial AI and (2) a control hearing aid representing the current state of the art in "AI hearing aids" – on intelligibility in challenging, multi-talker environments. This double-blinded, randomized controlled trial employed a crossover, within-subject design wherein 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.

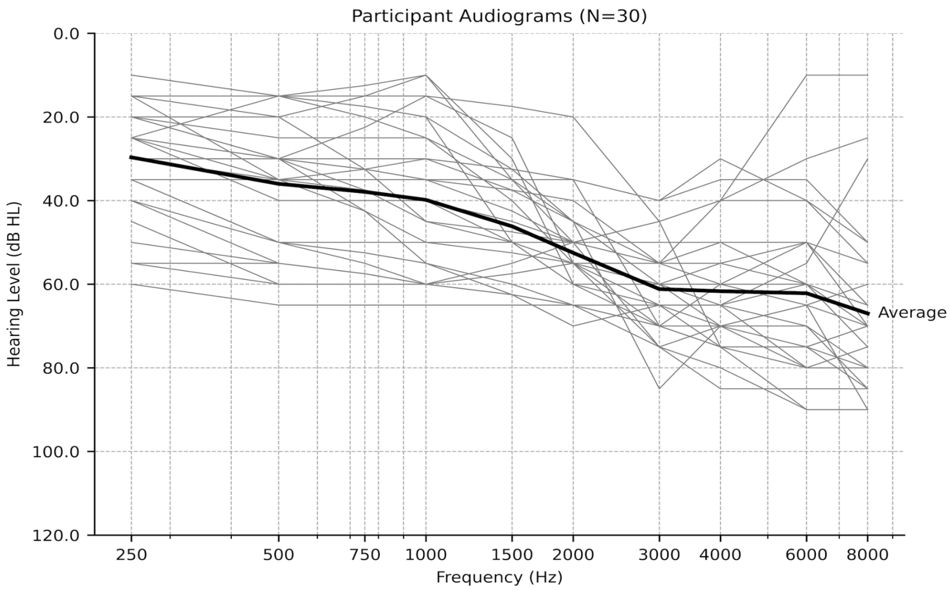

Figure 1: Participants' air-conduction audiograms (left ear shown for each participant).

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 HA was a receiver-in-canal hearing aid from a leading manufacturer. It was released in 2024 and represents the highest tech level for its product line. It employs real-time neural network-based noise reduction on a dedicated AI chip. This device was configured with a manual program using the manufacturer's fitting software such that both directionality and AI-based noise reduction were set to their highest levels. This device was selected because it represents state of the art technology.

The investigational HA used in the comparison was also a receiver-in-canal hearing aid. It runs the Spatial AI model and is trained to separate primary speech in front of the wearer (within an angle of ±60 degrees) from all other sounds, including non-speech noise from all directions and interfering speech from the back plane.

The model was trained using machine learning techniques. During training, the model's parameters were iteratively adjusted using backpropagation and stochastic gradient descent to minimize a loss function across many examples. The model was trained on both synthetic and real data. The synthetic data were created by mixing together clips of speech and noise. Both clean speech and noise training examples were created using a spatial sound simulator so that the model received spatialized audio and learned to exploit the timing differences between microphones. The real data exposed the model to realistic conditions (e.g. moving sources, head rotations, reverberant settings, etc). In training, the model predicted the clean speech in front of the wearer and model parameters were iteratively adjusted to improve the accuracy of the prediction. The model was trained over millions of audio samples representing years of audio data. Spatial AI was not trained using any of the test materials used in the study.

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

Both sets of hearing aids were programmed to NAL-NL2 fitting targets (National Acoustic Laboratories, Sydney, Australia) according to the audiogram so as to control for differences in audibility. All hearing aids were adjusted so that their gains were within 5 dB of target at 500, 1k, 2k, 4k Hz, using a Verifit 2 (Audioscan, Dorchester, Ontario, Canada) with the international speech test signal (ISTS) as the stimulus. All fittings were done in the General program (or its equivalent) using both manufacturer's fitting software. Both used power domes to occlude the ear and isolate the impact of the amplified path of sound.

Speech-in-Noise Testing

After the hearing aids were adjusted according to their audiogram, each participant completed Speech in Noise testing as the primary endpoint of interest. Testing was completed over 3 days in a recording studio in New York City. Data collection was overseen by one of the academic research collaborators, who was on site for testing the majority of the participants. The academic research collaborator was charged with verifying the integrity of double blinding and double checking data scoring.

Speech intelligibility was assessed using the QuickSIN test, a widely used measure of speech recognition in noise. 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, which consist of a first track with a target speaker saying complete sentences and a second track containing four simultaneous competing talkers ("noise"), were unaltered from the typical QuickSIN. 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).

Participants were randomly split into 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 an acoustically-treated room of approximately 20 square meters. Participants were positioned on a chair in the middle of a 4-loudspeaker array (Genelec 8010s, Genelec Oy, Iisalmi, Finland), as shown in Figure 2, with loudspeakers positioned at 0 degrees (the “front loudspeaker”), 135 degrees, 180 degrees and -135 degrees (the 3 “back loudspeakers”). Loudspeakers were each positioned 1.2 meters from the head of the participant at eye level.

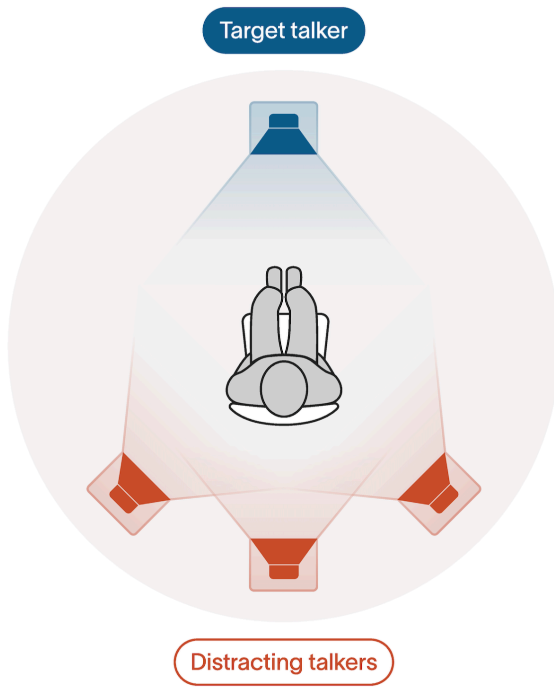

Figure 2: Diagram illustrating positioning of speech and noise sources relative to the participant. Noise (multi-talker babble) was played out of one of the three loudspeakers behind the participant. Which loudspeaker the noise played out of varied randomly with every sentence, though the test followed the same pattern of loudspeaker selection for every participant.

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.

Lab-measured Objective Metrics

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) with the investigational and control HAs. Both HAs were tuned to NAL-NL2 targets for the N3 audiogram and power domes were used to occlude the ear. 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 perform the Hagerman-Olofsson measurements, the antifeedback system, impulse noise reduction and wind-noise reduction was turned off on both devices.

Results

Intelligibility Testing

Scores were the percent of target words (5 target words per sentence) identified correctly. Figure 3 shows the percent correct for each hearing aid, with 95% confidence intervals for each SNR. At the most challenging SNR of -6 dB, the percent of words identified correctly 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.

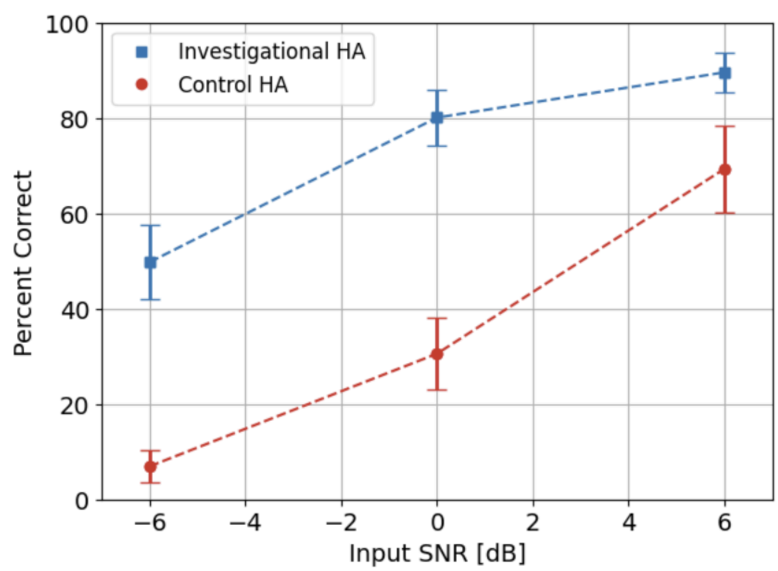

Figure 3: Average percent of words identified correctly for each hearing aid type for each SNR. Error bars show 95% confidence intervals.

We fitted the data using a generalized linear mixed-effects model with a logistic link function. This model treats which hearing aid is being used, the order of testing (i.e. which hearing aid was tested first) and the SNR as fixed effects, while treating the subjects as random effects. 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 via a Wald test on the 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 understanding words correctly (relative to the control HA) by 18.9x at the most challenging SNR, 10.6x at the middle SNR and 6.0x at the easier SNR. In all cases, the 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 model allowed us to derive the SNR-50 (the SNR at which participants correctly identified 50% of the words). Specifically, we plugged the model's fitted coefficients into its own equation, set the predicted probability to 0.5, and solved for SNR. The SNR-50 of the control HA was 3.08 (95% CI: [2.84, 3.31]) dB while the SNR-50 of the investigational HA was -6.15 (95% CI: [-6.86, -5.45]) dB, giving an SNR-50 improvement of 9.2 dB. These results match closely with what one gets from inferring the SNR-50 visually on Figure 3.

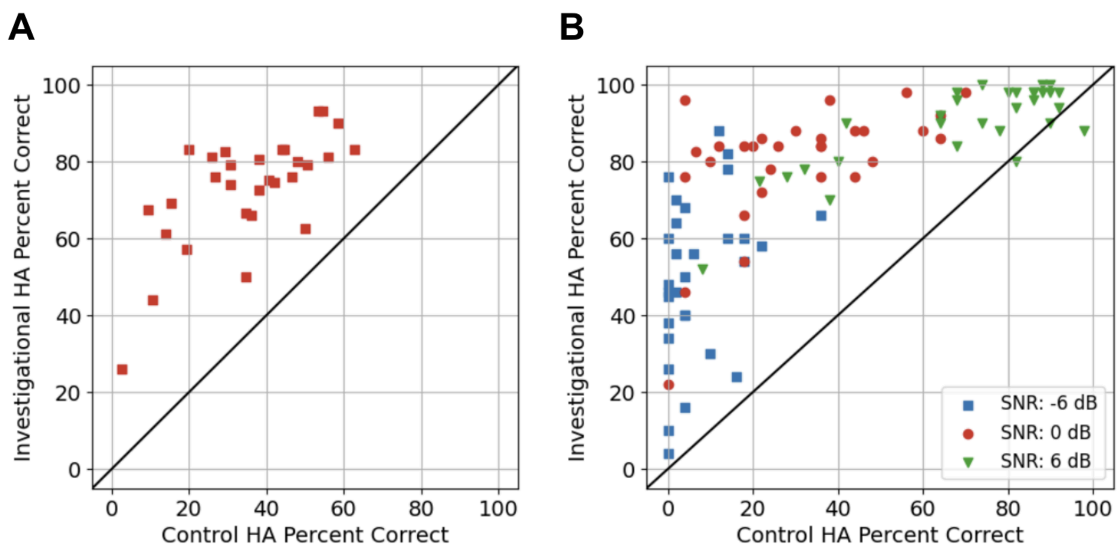

Figure 4: Panel A shows the percentage of words identified correctly wearing the investigational hearing aid and the percentage of words identified correctly wearing the control hearing aid for each participant averaged across all tested SNRs (n=30). Figure 4b shows the same data broken down by SNR (n=30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

Figures 4a and 4b show results for each participant. Figure 4a shows the percent words correct for each participant wearing the investigational hearing aid plotted against the percent correct wearing the control device (averaged across all three SNRs). Notably, every study participant did better wearing the investigational hearing aid, the improvement being statistically significant for every individual. Figure 4b shows the same performance 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.

Lab-measured Objective Metrics

The results of speech intelligibility testing with human subjects were compared to lab-measured objective metrics. We used the Hagerman-Olofsson phase inversion method (Hagerman et al. 2004) to estimate the SNR of the investigational and control HAs on the QuickSIN test. This technique allows separating the speech and noise components of a simultaneously-presented audio signal, and thus calculating the SNR. Let s and n correspond to the speech and noise components of the stimuli presented through the loudspeaker array. The Hagerman-Olofsson method involves making three KEMAR recordings:

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

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

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

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

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

The quantity

$$E = (m_\epsilon + m_o) / 2$$

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:

$$\epsilon_s = 10\, log_{10}(< E^2 > / < S^2 >)$$
$$\epsilon_n = 10\, log_{10}(< E^2 > / < N^2 >).$$

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

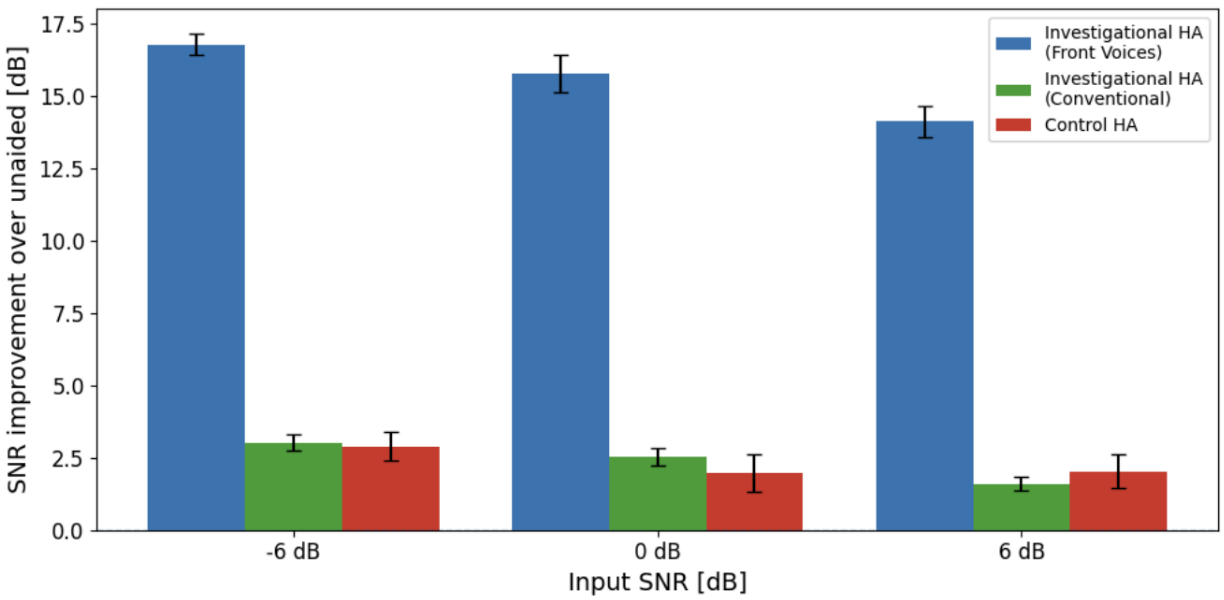

Figure 5: The SII-weighted SNR improvement (over the unaided condition) at the three input SNRs for the investigational HA (Front Voices mode, which uses Spatial AI), the control HA and Conventional program (no AI) on the investigational HA. Error bars show 95% confidence intervals.

Figure 5 shows the SII-weighted SNR improvement (over the unaided condition) for the investigational HA and the control HA. We also measured the Hagerman-Olofsson SNR using the "Conventional" program on the investigational hearing aid. This program does not use AI and includes the standard suite of hearing aid algorithms, including classical noise reduction, adaptive beamforming and wide dynamic range compression (WDRC). Averaging over all input SNRs, we found that the investigational HA (in Front Voices mode) achieved an SNR improvement of 15.5 dB, as compared to 2.3 dB for the control HA. The Conventional program achieved an SNR of 2.4 dB, nearly identical to the control hearing aid. In terms of standard SNR (not SII-weighted), the improvement over unaided was 17.1 dB, 1.9 dB and 2.6 dB for the investigational HA (Front Voices), control HA and the investigational HA (Conventional), respectively. 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.

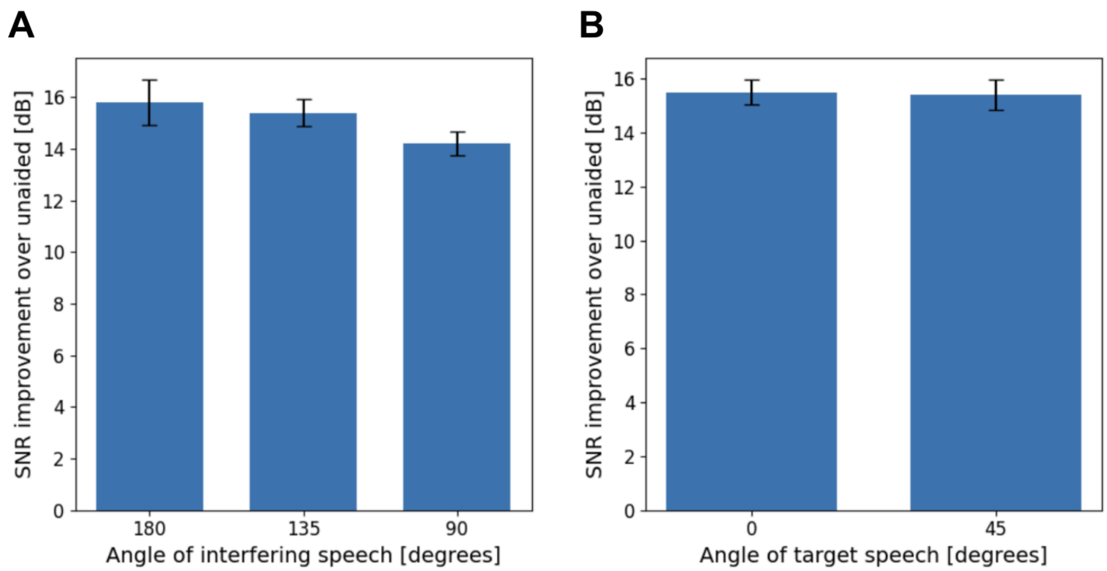

Figure 6: Panel A shows the SII-weighted SNR improvement (over the unaided condition) as the angle of the loudspeaker that plays out the interfering speech is varied (with target speech coming from the front). Panel B shows the SII-weighted SNR improvement (over the unaided condition) as the angle of 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 how the SNR improvement of the investigational HA varied as the loudspeaker playing the noise was moved through different angles. The SNR improvement was 14.2 dB with interfering speech coming from 90 degrees, only 1.6 dB lower than the SNR when the interfering speech was directly in the back. Figure 6b shows that the SNR improvement was 15.4 dB when the target speech was coming from 45 degrees relative to the front (with interfering speech in the standard configuration of ±135, 180 degrees), nearly identical to the SNR when the speech was directly in front.

We also computed the Hearing-Aid Speech Perception Index (HASPI) v2 (Kates et al. 2021), a metric intended to model speech intelligibility. The investigational HA scored 0.24, 0.57 and 0.77 at input SNRs of -6, 0 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.

Discussion

Our results indicate that Spatial AI processing confers a significant gain in speech intelligibility in multitalker noise for individuals with hearing loss. Furthermore, the control HA is representative of the most advanced signal processing commercially available as of the time of writing; as configured for the study, the control HA uses directionality and deep neural network-based noise reduction, with both functions set to their highest level in the fitting software.

An important question is why Spatial AI shows such a significant intelligibility improvement where neural network approaches available on today's hearing aids have not. One key difference for Spatial AI from many of these approaches is the spatial element itself, i.e. the fact that the model synthesizes information from multiple microphones on the hearing aid. This spatial information is used to sort desired sounds from undesired sounds in a way that is highly reliable and thus allows a significantly larger SNR improvement than single-microphone algorithms can achieve.

Another reason for the large effect size may be that the neural network is particularly well suited to handle noise sources that produce informational masking, since the model may not have the same susceptibilities. Multi-talker babble contributes more informational masking than, for example, steady noise, which contributes more modulation masking (Stone et al. 2012). The Spatial AI model, using the direction of arrival for different sound sources to sort desired from undesired speakers, may not be impeded by "informational masking" in the way that a human listener is.

Spatial AI processing ameliorates a major frustration for hearing aid wearers by improving the SNR in multi-talker environments. Furthermore, the 9.2 dB improvement in SNR-50 is large enough to motivate hearing aid wearers to adopt the new technology. The SNR improvement that is meaningful to hearing aid wearers has been explored using an empirical approach (McShefferty et al. 2015). They found that an SNR improvement of 3 dB was "noticeable" and a 6 dB improvement would be large enough to motivate action, in this case by visiting the clinic. The SNR-50 improvement in the study is well above the threshold for action.

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.

These findings imply the approach of implementing machine learning models on dedicated processing chips is yielding rapid advance. The superior performance of the investigational hearing aid over an already-advanced, modern, AI-powered hearing aid suggests that this technological paradigm likely represents the beginning rather than the culmination of improvements in hearing aid signal processing.

Limitations and Future Research

One limitation of this research is that the study focused only on speech in multi-talker noise. Future research will examine other noise types, or mixes of noise types. It also only tested positions to the rear of the wearer for the competing noise source, and in a room with minimal reverberations. Further research may look at other spatial configurations, with noise coming from different directions and different angles, and in other environments. It should be noted that the Front Voices program on the investigational HA is designed to be manually engaged by the user when the voices of interest are in front. If a user engages this program when trying to have a conversation with a speaker at their side (for example when going on a walk), it could reduce rather than improve their intelligibility. Similarly, if an interfering speaker is within 60 degrees from the front, the Spatial AI model will not recognize them as an interferer and they will be preserved in the output. Additionally, the choice of power domes, which minimizes the effect of the direct path, is unlikely to be replicated by wearers who are optimizing for all-day wear, so further research may assess whether the benefit decreases with other types of domes. This study only used three challenging SNRs, and even for the most challenging SNR, intelligibility with the investigational HA was almost 50% on average. It would therefore be instructive to test even more challenging SNRs. 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 are widely understood to be the most challenging situations because other speech is the most distracting noise source and hearing aids are generally designed to amplify speech. Hearing aid wearers frequently complain that they can hear the conversation at another table more easily than they can hear the person across from them. With Spatial AI processing, we see evidence of a real-time system that led to significant intelligibility improvements for hearing aid wearers in such scenarios: 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.

As machine learning models become more sophisticated and dedicated processing hardware continues to advance, it appears there is considerable potential for even greater improvements in speech intelligibility—particularly in the challenging "cocktail party" environments that most significantly affect quality of life for hearing aid users.

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.