Spatial AI Consistently Preferred to State of the Art Hearing Aids in Multitalker Noise

Cole Morris[1,5], Igor Lovchinsky[1,5], Christina Callahan[1], Kathleen Wallace[1], Mark Berry[1], Kaitlyn Lyngaas[1], Israel Malkin[1], Nathan Agmon[1], Bess Nagler[2], Joseph Montano[3], Barbara Weinstein[4]

[1] Fortell Research
[2] Mount Sinai Hospital
[3] Weill Cornell Medicine
[4] Graduate Center CUNY and NYU Langone Medical
[5] Co-first authors/contributed equally to this work

Abstract

Objective
We examined a spatial AI model running in an investigational hearing aid to see whether it improves perceived ease of understanding speech in noisy, multitalker environments relative to hearing aids using more traditional AI models. The spatial AI model is a novel neural network algorithm that uses information from multiple microphones to isolate the voices in front of the wearer.

Design
In a blinded, randomized-controlled trial, participants compared the investigational hearing aid to five control hearing aids, representing the latest technology from the five major hearing aid manufacturers. Recordings of the hearing aids in their speech-in-noise programs were made on KEMAR in noisy, multitalker environments. In each of 100 randomized A/B trials, participants rated which hearing aid made a target talker easier to understand.

Sample:
Twenty adults with mild to moderately-severe bilateral sensorineural hearing loss.

Results
The investigational device was preferred to the control hearing aids in 99 out of 100 comparisons, with one tie.

Conclusions

For most clinicians, performance in noise is a top consideration when selecting a hearing device. The consistent preference for the investigational hearing aid suggests that spatial AI may significantly improve perceived ease of understanding in challenging, multitalker noise.

Keywords:

speech-in-noise, hearing aids, machine learning, deep neural network, speech enhancement, listening effort

Introduction

The most challenging conversations for individuals with hearing loss are conversations in the presence of loud multitalker noise. In these environments, the signal-to-noise ratio (SNR) can be very difficult and the type of noise– other voices– is maximally challenging: loud in overlapping frequencies, non-stationary and distracting, resulting in informational masking. Individuals with hearing loss struggle to focus on the voice of their communication partner as they attempt to separate the signal from the noise.

Hearing aids use a variety of signal processing techniques to improve listening in these challenging environments, all with the goal of improving the signal-to-noise ratio and ease of understanding without degrading perceived sound quality. Most hearing aids on the market today employ multiple algorithms simultaneously including beamforming, stationary noise reduction and program adjustments that fine tune the hearing aid processing for different environments. All of these techniques incrementally improve the listening experience in noise. However, in multitalker scenarios, the canonical “cocktail party” scenario, the same dynamics that challenge the hearing impaired listener simultaneously challenge traditional hearing aid algorithms. In these situations, much of the noise is non-stationary and overlaps with the target speech in both amplitude and frequency. As such, traditional noise reduction techniques can only remove so much noise before simultaneously removing or distorting necessary components of speech. Furthermore, most traditional algorithms are fundamentally designed to enhance speech and thus most struggle when the competing noise is other people talking. Of the techniques mentioned above, beamforming is the most useful in multitalker scenarios, but frequently insufficient.

More recently, machine learning techniques, in particular deep neural networks, have offered a new approach to noise reduction. These algorithms can treat sounds differently based on what they are at a semantic level and thus are better suited to navigate the complexities of real-world situations than traditional signal processing techniques, which rely on heuristics to separate speech and noise. But these AI algorithms present other challenges. The first is that most hearing aids are highly constrained devices, needing to hide behind the ear and run all day on a tiny battery. This limits the size of the neural network algorithm that can be run, which in turn limits its performance. The second challenge is that these AI models are typically trained to pick up on speech, so they still frequently amplify the noise when that noise is distracting other talkers.

Across both traditional signal processing and machine learning techniques, numerous studies have attempted to examine whether and to what degree different signal processing approaches appreciably improve the listening experience for hearing aid wearers. Bentler et al. (2005) provided an early systematic review of digital signal processing techniques and found that directional noise reduction techniques were effective but digital noise reduction techniques had scant evidence of effectiveness. Since then, there have been significant advancements in hearing aid technology, but the evidence of effectiveness is generally mixed. Johnson et al. (2016) and Cox et al. (2016) compared premium and basic hearing aids from multiple manufacturers in both laboratory and real world environments. In the lab, only one of four brands saw improved listening effort from turning on premium features, and none saw improvements in either intelligibility or longitudinal outcomes. Wu et al. (2019), also comparing premium to basic hearing aids, found that premium features did improve speech reception thresholds by a statistically significant 1-2 dB, but not by enough to yield lower perceived listening effort or improved sound quality ratings. Other studies geared particularly at more advanced directional microphone technologies (e.g. binaural beamforming or other directional noise reduction techniques) have shown some impact on listening effort in the lab (Picou et al. 2017; Dong et al. 2024; Valderamma, 2025) but these benefits are often modest in size, vary across test conditions, and rarely translate into large, consistent improvements in everyday listening for most users. Comparative, blinded evaluations across multiple premium devices are uncommon, and the few that exist suggest that the perceptual gap between current top-tier products is typically narrow.

The purpose of this study is to compare a form of novel spatial AI processing running in an Investigational Hearing Aid to the state of the art in hearing aid technology in a blinded randomized-controlled trial. Spatial AI is a neural network algorithm that exploits the timing differences of audio between microphones to make inferences about the location of sound sources. The model is trained to amplify voices in front of the wearer and remove non-speech noise from all directions and speech behind the wearer.

For most clinicians, as well as potential hearing aid wearers, performance in noise is a top consideration when selecting a hearing device. That said, differences between offerings are sometimes hard to hear, and most comparative listening is poorly controlled or rarely takes place in clinical practice. Therefore, a carefully designed blinded listening experiment provides useful information for an audiologist seeking to identify the best technology for prospective patients. For this study, we used KEMAR to record the output of six different hearing aids (the Investigational Hearing Aid and the latest hearing aid from each major hearing aid manufacturer) in realistic “cocktail party” noisy scenes. We then had subjects with hearing loss select which hearing aid, operating in their default speech-in-noise modes, made it easier to understand a target speaker in these environments. This provided a robust, unbiased dataset to examine whether spatial AI would improve hearing aid wearers' perceived ease of understanding speech in multi-speaker noise.

Materials and Methods

Participants

20 adults with mild to moderately-severe bilateral sensorineural hearing loss were recruited for the study. 18 of 20 were hearing aid wearers and the other two had been assessed as hearing aid candidates. All participants provided written informed consent prior to participation and were compensated for their time. The study was approved by WCG Clinical IRB (IRB Study #: 1394583).

Participants provided audiograms taken within the last 16 months or had the option to receive an audiogram by a licensed audiologist as part of the study. The inclusion criteria for the study were as follows: a. participants had a four frequency pure-tone average greater than 20dHL and less than 70dBHL, as verified by the submitted audiogram (they also confirmed no perceived changes to their hearing since their audiogram); b. participants spoke English as a primary language and c. Participants owned a personal computer and headphones. Individuals were excluded if they had a known cognitive impairment or had asymmetry between ears of greater than 20 dB based on the four frequency PTA. Note that compensation for participants' hearing loss was facilitated by matching participants to one of four standard audiograms that the survey supported. If participants' audiograms diverged too much from the standard audiograms supported, they were excluded from the study. This method is described in full below in the section on 'Hearing Aid Fitting'.

All participants completed the Hearing Handicap Inventory for the Elderly- Screening HHIE-S (Ventry and Weinstein, 1982) with a researcher before participating. Responses of experienced hearing aid users were based on how participants communicate with their current hearing aids. Table 1 provides summary data for the 20 participants' ages, pure tone averages and HHIE-S scores.

Table 1. Participant Hearing Evaluation Data (N=20)

	Min	20%	50%	80%	Max
Age	38	62	69	80	87

Pure Tone Average Air conduction, both ears Avg (500, 1k, 2k, 4k), dB HL	31	42	46	56	66
HHIE-S	6	12	19	29	32

Listening Survey

The study was designed as a listening survey taken on the computer, where participants could listen to pairs of recordings of two different hearing aids in close succession. Using the online survey platform, participants listened to 100 A/B comparisons, each of which compared the Investigational Hearing Aid running Spatial AI to one of 5 Control hearing aids (20 questions each).

Each question prompted the listener to listen to a “cue clip”, containing 4 seconds of the target talker speaking, and then listen to two recordings (Recording A and Recording B) of a sound scene containing that target talker. Both recordings A and B were of the same 6 seconds of audio; one of the recordings was of the investigational hearing aid while the other was of one of the five control hearing aids. The cue clip was of a different 4 seconds of the target talker speaking. The instructions for each question were the same: “In which sample is it easier to understand what the target speaker is saying?” Participants then provided a response on a 5 point scale:

1) A much easier
2) A slightly easier
3) Both samples similar
4) B slightly easier
5) B much easier

The survey was administered via an online survey platform developed by the researchers. The study format of an online survey was selected because listening to recordings of the devices allows immediate, seamless switching between the audio of two different hearing aids. Using the physical devices requires the physical device to be switched with every comparison. The inherent delay in physically switching devices both impedes the ability of the listener to make a high quality comparison and the time involved in making the switch makes it significantly impractical to gather a high number of high-quality, randomized, blinded comparisons between hearing aids. Also, by making recordings of each hearing aid under tightly controlled conditions, we were able to minimize the influence of other variables that would normally impact subjective listening experience (variation in physical fit, physical positioning in the sound field, delays between hearing recordings).

Hearing Aids

This study compared the Investigational Hearing Aid running Spatial AI to 5 Control Hearing Aids.

The Control hearing aids were all behind-the-ear, receiver-in-canal hearing aids from the five major manufacturers of hearing aids; the highest level of technology was selected to represent each. All hearing aids were released in either 2024 or 2025 and advertise AI features ranging from AI based environmental classifiers to AI-based real-time noise reduction algorithms. The brand and model of each was selected by our academic research collaborators to represent, in their opinion, the best commercially available hearing aid for each manufacturer.

The Investigational Hearing Aid used in the comparison is also behind-the-ear, receiver-in-canal hearing aid, which runs the Spatial AI model. As described before, the Spatial AI model is trained to recover speech originating from in front of the wearer, using supervised learning techniques applied across millions of audio examples. Training data includes both real-world recordings and synthetic mixtures created by combining clean speech with noise. Some of thetraining data is spatialized using a sound simulator, allowing the model to learn to infer the direction of arrival for different sound sources based on the timing differences of the signals between microphones.

Model parameters are optimized using backpropagation and stochastic gradient descent to minimize a loss function over many iterations. During training, the model gradually improves its prediction of the desired clean speech signal in each noisy example. Altogether, the training corpus represents the equivalent of several years of audio data.

Because the resulting model exceeds the computational limits of conventional hearing aid DSPs, the device includes a custom-built machine learning coprocessor designed specifically to run this network. The processor is capable of executing up to 100 billion operations per second, allowing the model to run continuously without impacting battery life or device form factor.

Recording Setup

For the test stimulus, we created 100 spatially-realistic multi-talker sound scenes, each 6 seconds long. Each sound scene was composed by mixing together multiple sources:

1) **Background noise** was selected from the Ambisonics Recordings of Typical Environments (ARTE) Database (Bucholz & Weisser, 2019), which provides spatially realistic recordings of real world environments. We used the 7 loudest scenes (Cafe 1, Cafe 2, Dinner Party, Street Balcony, Train Station, Food Court 1 and Food Court 2).
2) **A target talker** was selected from the VoxCeleb2 dataset (Chung et al. 2018), which consists of thousands of interviews with different individuals. The advantage of the

VoxCeleb2 dataset is that it is highly representative of real conversational speech and spans a wide variety of voices, including accents, emotional states and conversational dynamics. 56 unique speakers (32 male and 24 female) were represented in the survey recordings.

3) **3 competing talkers** were also selected as noise from the VoxCeleb2 dataset.

None of the test materials were used to train the Spatial AI model in the investigational hearing aid.

To create the materials for the survey, recordings were made in a recording studio in New York City. 8 Genelec 8010 speakers were positioned uniformly around a circle with a radius of 1 meter. KEMAR was positioned in the middle, facing a front speaker at 0 degrees. Figure 1 shows how the sound sources were positioned relative to KEMAR. The background noise, which were ambisonic recordings, were played out of all eight speakers. The target talker played out of the speaker in front of KEMAR, positioned at 0 degrees. The three competing talkers played out of the three speakers positioned behind KEMAR at 135, 180 and -135 degrees.

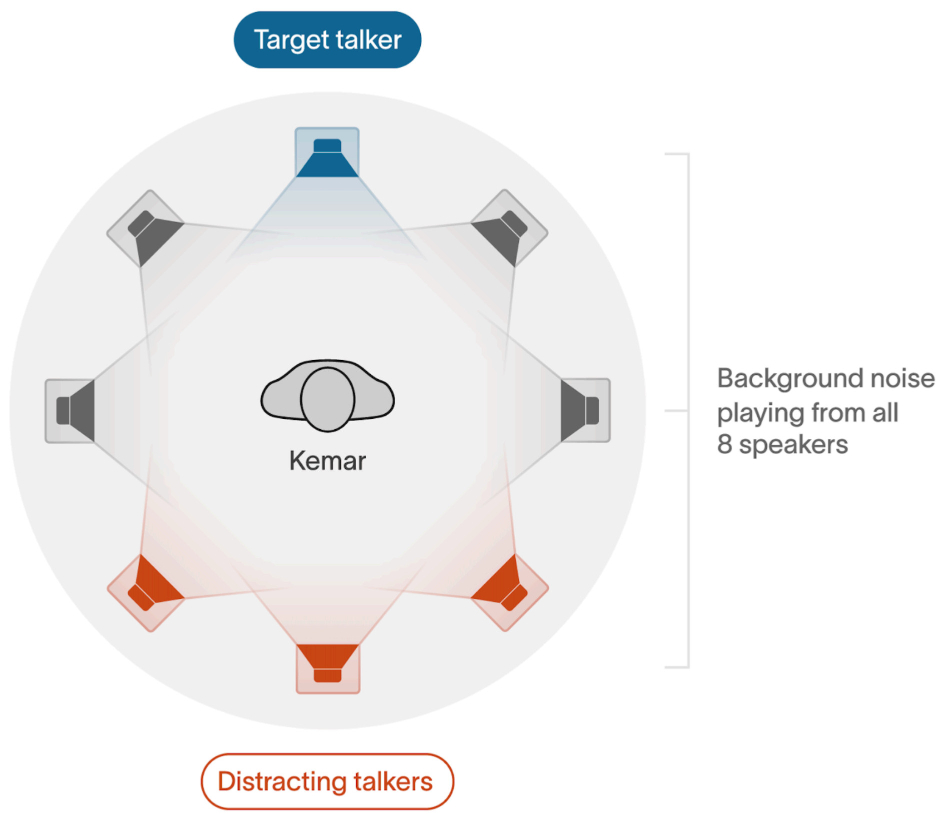

Figure 1: This diagram shows the layout of the speakers positioned around KEMAR and what types of sound were played from each the speakers. KEMAR is positioned facing the Target talker at the top of the circle.

To make the recordings, each pair of hearing aids were positioned on KEMAR's ears. The same power domes were used for all hearing aids to ensure an equally occlusive fit. In order to ensure that all hearing aid algorithms, many of which adjust to noise levels over longer time periods, were functioning to their fullest capacity, all sound scenes were recorded in a continuous noisy recording and the first several scenes were discarded. During the survey, the presentation volumes for the two recordings being compared in each question were equalized to ensure that comparisons were on the basis of sound quality and signal-to-noise ratio, since volume differences can easily be changed by an audiologist fine tuning the fitting or by the user adjusting the volume themselves.

Across the 100 different sound scenes, free field SNRs averaged -4 dB with a standard deviation of 2.8 dB. Since all hearing aids used beamforming, the SNR post-beamforming was centered at approximately 0 dB.

Hearing Aid Fitting

In order to adjust for the hearing loss of the participants, we created four versions of the test, each version corresponding to a standard audiogram. When a potential participant was recruited for the study, we checked whether their audiogram was a close enough match to one of the standard audiograms according to our predetermined matching criteria (described below). Participants were only enrolled if they had a matching standard audiogram that allowed for sufficiently accurate compensation for their hearing loss. If they did not match to a standard audiogram within our matching criteria, they were excluded. Participants then completed the version of the study corresponding to their matching standard audiogram.

The standard audiograms and matching criteria were either borrowed directly from or adapted from previous work, in particular the standard audiograms created for the international standard IEC 60118-15 (Bisgaard, 2010). Bisgaard et al. determined ten "standard audiograms" that covered the full range of hearing losses from mild to profound hearing loss. Seven of these are for flat and gradually sloping audiograms (N1-N7) and three are for steeply sloping hearing losses (S1-S3). For the study, we selected the three most common audiograms from the flat to gradually sloping group (N2-N4) and created one sloping audiogram (which we will call "S-mod") to best capture the S-group with a single audiogram. (Bisgaard et al. built on work from the Nordic Working Group which had also attempted to cover a range of hearing losses with standard audiograms. They had limited themselves to 5 and ended up with a single sloping audiogram, which is quite close to the S-mod audiogram we selected).

Similarly inspired by the approach taken in Bisgaard et al. (2010), we established a matching criteria whereby a participant audiogram could be considered a match to our standard audiograms. Participant audiograms were matched based on Euclidean distance to the closest standard audiogram, at which point the matching criteria was applied. We considered 4 key frequencies (500, 1000, 2000 and 4000 Hz). Audiograms were considered a match if all key frequencies were within 10 dB of the standard audiogram on both ears, with one deviation (>10 dB, <= 20 db) allowed per ear. Any deviation greater than 20 db invalidated a match.

Figure 2 shows the standard audiograms used for the study and all the matching hearing losses, with each dotted line corresponding to one ear of the participant (both ears were required to match). There were 3, 9, 5, and 3 participants for the N2, N3, N4 and S audiograms respectively.

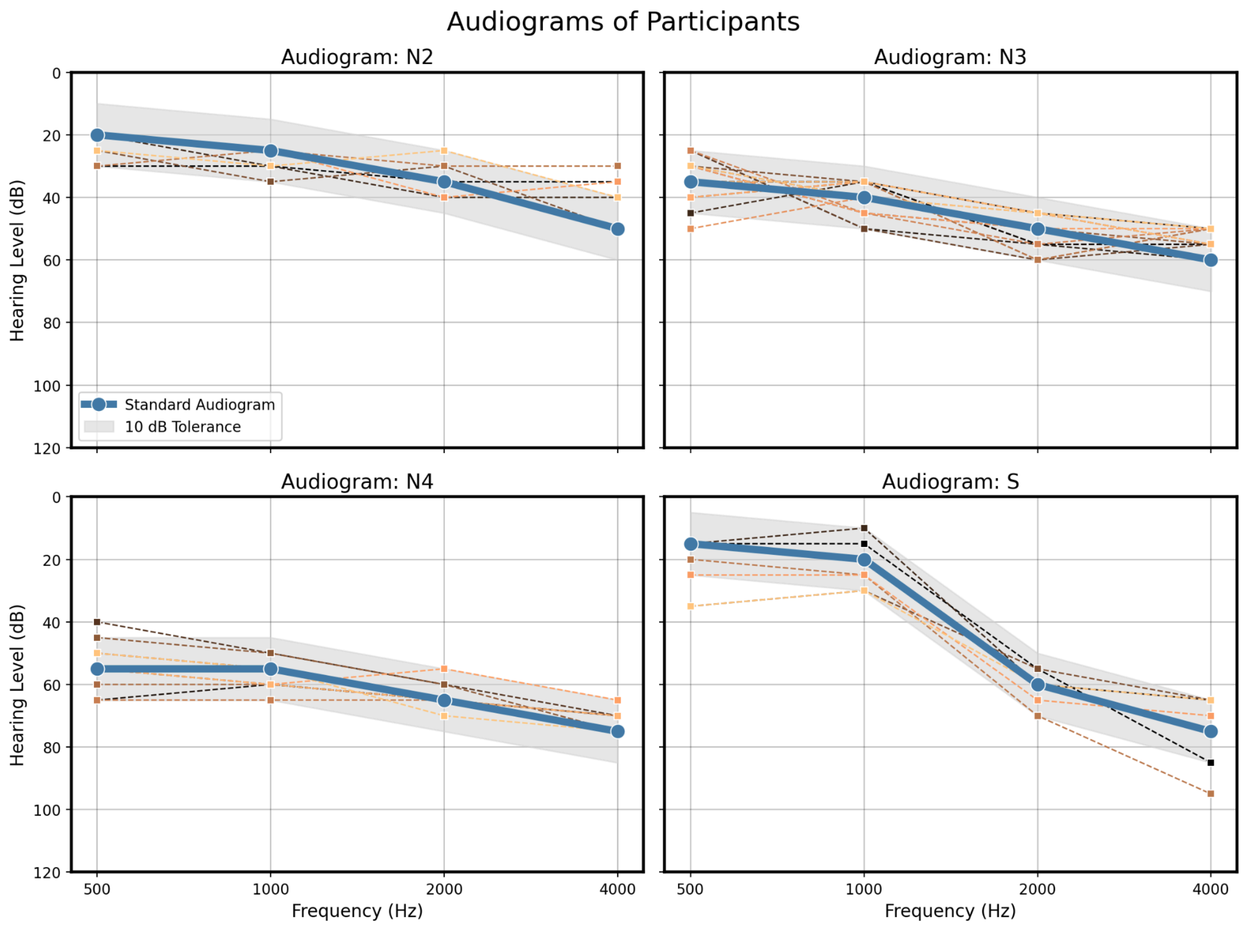

Figure 2 shows the 4 Standard audiograms in blue and the participant audiograms (both ears shown) that were matched to each standard audiogram. The grey shading shows the range where an audiogram was considered a match, with 1 point being able to diverge beyond the shaded area up to 20 dB.

For each standard audiogram, all hearing aids used in the study (the Investigational HA and the 5 Control HAs) were adjusted to the NAL-NL2 targets for that audiogram in a test box using a Verifit 2. All fittings were done in the General program (or its equivalent) using the manufacturer's proprietary fitting software. For all hearing aids, a Speech-in-Noise program was added so that it could be selected manually. Default signal processing settings were used for the Speech in Noise program in all cases. Many of the hearing aids use AI features in their Speech in Noise program by default and these were always left on. If there was a "Speech in Loud Noise" option, that was activated. Additionally, programs were linked to the General program so that fine tuning adjustments made to hit NAL-NL2 targets would carry over to the Speech in Noise program. During testing, all hearing aids were put in the Speech in Noise program manually.

Procedures

All participants met with a researcher via Zoom before completing the survey. During the Zoom call, informed consent was collected and participants were then provided with instructions for completing the survey. Participants also completed an example question (not part of the 100 survey questions) while on the Zoom to confirm that they understood the instructions. They were not provided any feedback on their response to the example question and that question was excluded from the analysis. After the Zoom call, the researcher sent the participant a link to the survey (the link corresponding to a particular version of the survey that compensated for their hearing loss using their matching standard audiogram) and participants completed it on their own time. Based on logs in the survey platform, the median time to complete the survey was 52 minutes.

Randomization and Blinding

To remove any potential bias for a given participant, whether the investigational hearing aid was recording A or B, which sound scenes were assigned to each of the control hearing aids (20 questions each), and the order of the questions were all randomized. Furthermore, we applied additional steps to minimize systematic errors across the participant sample. First, for every participant who completed the survey, a second participant completed a “flipped” version of the survey where the questions were otherwise the same but the position of A and B were flipped (we allowed participants to play recordings more than once, but were concerned that position B might be systematically favored because participants would listen to it second, when they were more familiar with the noisy mixture). Second, we created 5 different permutations of the survey, rotating which 20 sound scenes were assigned to which control hearing aid so that each control hearing aid was equally exposed to all the recordings. Participants were randomly assigned to a permutation with balanced assignment. Combining all of these methodological decisions, the maximum number of participants (total N = 20) that completed the same version of the survey was 2.

To ensure blinding, participants were not provided any context around what the recordings represented (i.e. no one knew that these were recordings of different hearing aids). Instead, participants were told that we were comparing different types of signal processing in hearing aids.

Results

Each participant's survey provided preference data for five head-to-head comparisons of the investigational hearing aid versus each of the five control hearing aids. To score each comparison, first we assigned values to responses based on the strength of the preference:

- Control much easier = -2
- Control slightly easier = -1
- No preference = 0
- Investigational slightly easier = 1
- Investigational much easier = 2

We then scored each participants' head-to-head comparisons by taking an average preference score for each question featuring that control hearing aid (20 questions per participant per control). If the average score is greater than 0, we considered that a preference for the investigational hearing aid. If the average score is below 0 we considered that a preference for the control hearing aid. If it's exactly 0, then we considered the results to show "No preference". Figure 3 shows which hearing aid was preferred in all 100 of the comparisons (20 participants x 5 control hearing aids).

Results from the blinded listening experiment showed an overwhelming preference for the investigational hearing aid. Participants preferred the Investigational device in 99 comparisons and had no preference in 1 comparison.

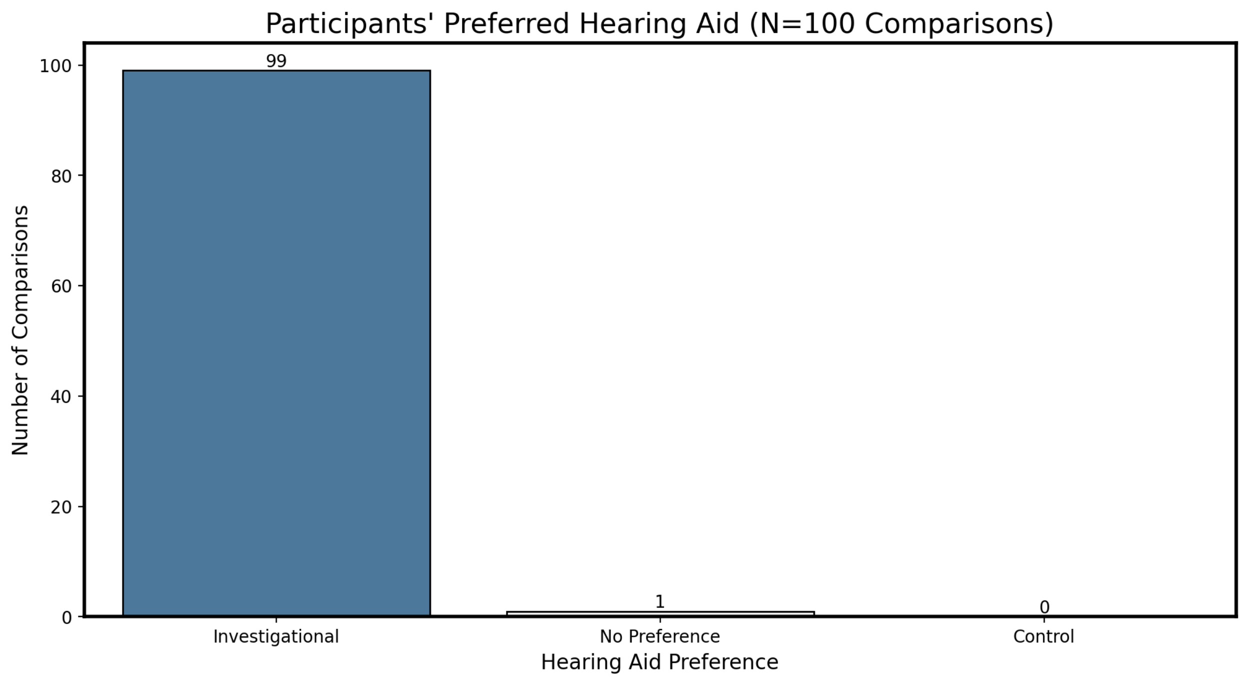

Figure 3 shows the preferred hearing aid for each comparison. There were 100 total comparisons (20 participants x 5 control hearing aids).

Figure 4 shows participants' responses aggregated for each of the 5 control hearing aids. The average preference score across all five control hearing aids was 1.15, with the highest average preference score of 1.29 and the lowest average preference score of 0.99.

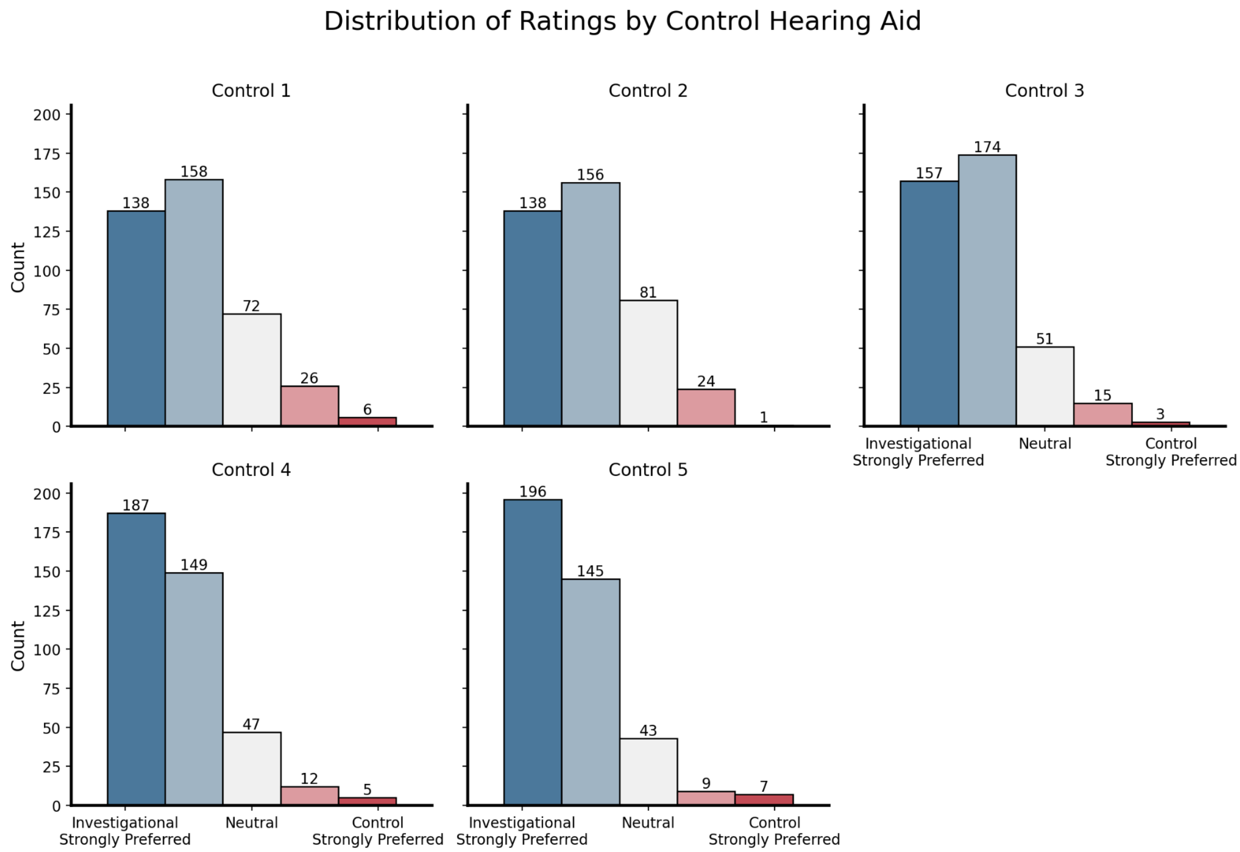

Figure 4 shows the aggregated responses across participants for each control hearing aid.

To estimate the average preference score for each control hearing aid, we used a linear mixed-effects model with random effects for participants and scenes and fixed effects for each control. This approach extends ordinary least squares regression by incorporating random effects and accounting for correlated errors. In this context, the model accounts for repeated measurements within participants and allows each participant to have their own baseline preference for the treatment hearing aid (i.e., a random intercept).

The model can be expressed as:

$$Y_{ij} = \sum_{k=1}^{K} \beta_k X_{ijk} + u_i + v_j + \varepsilon_{ij}$$

where:

- i, j and k are indices that refer to the participant, the scene and the control hearing aid, respectively

- Y_{ij} is the preference score for participant i on scene j
- β_k is the fixed effect for hearing aid k
- X_{ijk} is the indicator variable for whether participant i in scene j used hearing aid k
- u_i is the random intercept for participant i, where $u_i \sim N\left(0, \sigma_u^2\right)$
- v_i is the random intercept for participant i, where $v_i \sim N\left(0, \sigma_v^2\right)$
- ε_{ij} is the residual error, where $\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right)$

The investigational hearing aid was preferred over each control with high statistical significance ($p << 0.001$). Table 2 shows the average preference score by control hearing aid with 95% confidence intervals.

Table 2: Average Preference Score by Control Hearing Aid

Hearing Aid	**Preference Score**	**95% Confidence Interval**
Control 1	0.99	[0.90, 1.08]
Control 2	1.015	[0.97, 1.06]
Control 3	1.167	[1.08, 1.25]
Control 4	1.252	[1.17, 1.33]
Control 5	1.285	[1.20, 1.37]

Note: Preference scores are estimates from a mixed effects model with random rater intercepts. All results are statistically significant ($p < 0.001$).

The scores also show some variation in preference between control hearing aids. Because all hearing aids were compared to the investigational hearing aid, the scores are in absolute terms and can be used to determine whether certain hearing aids were preferred to others. There are some statistically significant differences between the best and worst control hearing aids, but the top to bottom spread in scores is much smaller than preference for Spatial AI relative to even the best of the controls.

Looking across participants, 19 participants preferred the investigational device relative to the controls in all cases, while 1 participant preferred the investigational device in 4 of 5 comparisons and had no preference in the 5th comparison. Notably, as shown in Figure 5 below, the strength of the preference depended on the participant and the control hearing aid being used.

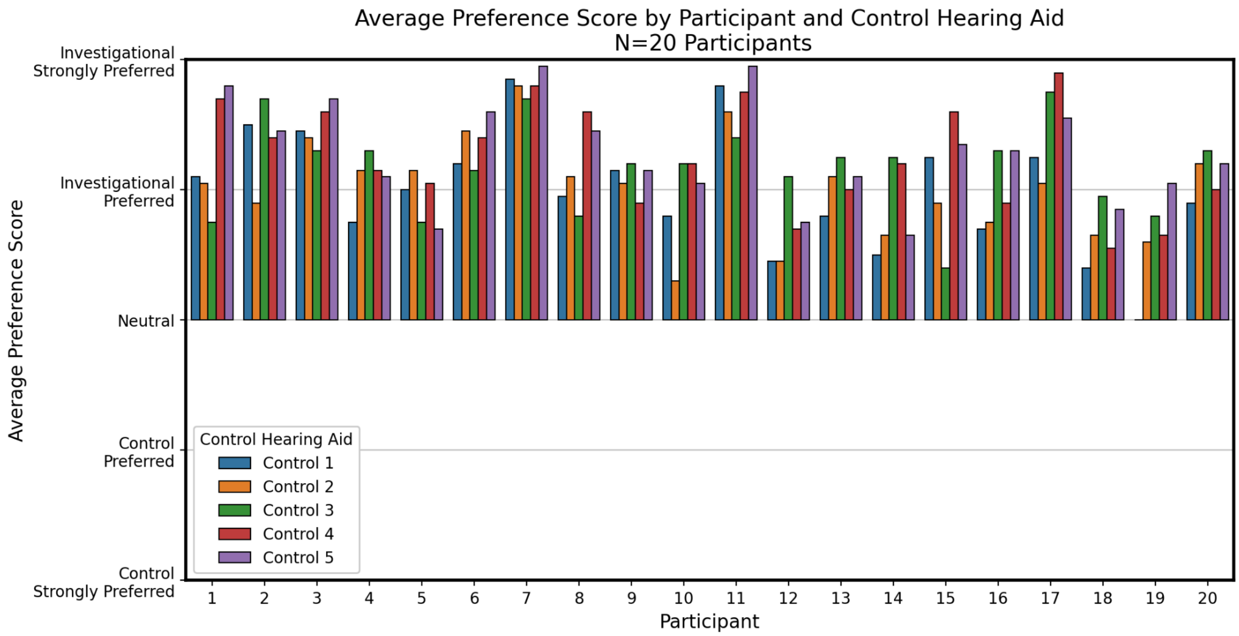

Figure 5 shows each participant's preference scores for the investigational hearing aid versus each of the 5 control hearing aids

One question relevant for audiologists will be whether there are potential configurations of hearing loss that predict where spatial AI will make the largest difference. To answer this question, we analyzed participants' average preference scores for the investigational device as a function of the standard audiogram to which participants were matched and their hearing handicap inventory (HHIE-S) score. We observed that the preference score for Spatial AI was similar across standard audiograms, suggesting that the appeal of spatial AI is not necessarily a function of the severity or configuration of the hearing loss. None of the differences between standard audiograms were statistically significant. Strength of preference showed a modestly declining relationship with HHIE-S scores ($R^2 = .29$), though the overall preference remained significant for both high and low handicap groups.

Discussion

Conversation in noise is the most common hearing difficulty among adults (ASHA, 2021) and yet it is the most frequently cited point of dissatisfaction with current hearing aid technology (Picou, 2022). It is notable that of the 18 study participants who wore hearing aids, only one had an HHIE-S score indicating no hearing handicap with their hearing aids; 66% reported a mild to moderate handicap and 26% still live with a severe handicap. Of course, some of this is the nature of hearing loss, but it also speaks to the persistent limitations of today's technology. Today's hearing aids have incorporated many signal processing advancements in handling speech in noise, yet 88% of study participants who wear hearing aids still reported that they struggle in restaurants. The consistent and strong preference for the Spatial AI processing relative to other technologies to which we compared the investigational device indicates that spatial AI may have a significant impact on the overall experience for hearing aid wearers, in particular in the noisy scenarios wearers find most challenging.

Until recently, most academic and clinical research has found that, when hearing aids are appropriately fitted, there are few consistent or perceptible differences in user experience between brands or between basic and premium technology tiers—especially in real-world conditions. Several blinded field studies, including those by Cox et al. (2016) and Wu et al. (2019), have shown that differences in speech intelligibility, sound quality, and user satisfaction between devices tend to be small, inconsistent, or absent altogether. This has led to a widespread view that, while manufacturers may implement signal processing differently, the end results are broadly equivalent for most users.

Results from our blinded listening experiment are broadly consistent with this consensus in the case of currently available hearing aid technologies. While we observed statistically significant differences between the control devices, they were modest in magnitude: the average preference scores for the highest- and lowest-rated control hearing aids differed by just 0.3 points on a 5-point scale. By contrast, the investigational device incorporating spatial AI was consistently and strongly preferred over all other devices, with a preference score difference to the best-scoring control approximately three times larger than the spread between the best and worst control devices. This result suggests that spatial AI may represent a more substantial perceptual advance in noise processing than any of the incremental innovations embedded in today's premium hearing aids.

Comparative listening in challenging, noisy scenarios is rarely part of the hearing aid selection process for most audiologists; this is likely in part because differences between brands are hard to discern. However, the growing diversity and opacity of AI-based features across manufacturers may make such comparisons increasingly important. All six hearing aids in the study have incorporated AI at least to some degree, and yet only one of the hearing aids stood out from the others in the blinded comparisons. If algorithmic innovation is outpacing clinician or user ability to understand it through specification sheets or marketing claims, perceptual evaluation—especially under ecologically valid noisy conditions—may offer a grounded way to identify meaningful differences.

Artificial intelligence presents a new frontier in signal processing. Given the nascent state of the science, it is likely that new algorithms, enabled by new hardware, will allow for larger and more significant differences between products to emerge. In this regime, studies like the blinded listening experiment done here, or alternatively, introducing blinded A/B comparisons directly to consumers, may allow hearing aid buyers to better understand which AI capabilities actually matter to the listening experience and which are simply marketing.

Limitations and Future Research

One limitation of this study is that it focused on the specific scenario of multi-speaker noise. Future research should examine relative preferences in other types of noise. Another limitation is that the approach for adjusting for people's hearing loss– matching to the standard audiograms– covered a wide range of standard hearing losses but did exclude individuals with more atypical hearing losses (asymmetries, cookie-bites, reverse slopes, etc). Future research should confirm that the preferences observed here generalize beyond the standard audiograms tested.

Furthermore, listening in loud, multi-talker scenarios is only a part of a hearing aid wearer's experience, so future research should explore whether the preferences observed in the lab translates to the same preferences in blinded field trials where users experience a variety of environments. Similarly future research should explore whether spatial AI yields differential improvements in measurable patient outcomes, like the HHIE-S scores measured in this study.

Lastly, as has been seen in other studies, perceived ease of understanding does not necessarily lead to objective improvements in intelligibility (and vice versa, as shown by Wu et al. (2019)). While this study did not measure participants' ability to correctly identify the target speech, objective intelligibility impacts of spatial AI were examined in a separate study, which showed that it improved SNR-50 for individuals with hearing loss by 9.2 dB relative to the top of the line AI hearing aid (Morris et al. submitted 2025, in peer review). Those findings, which also focused on multitalker noise at similar SNRs, suggest that the perceived ease of understanding observed in this study may extrapolate to objective improvements in intelligibility.

Conclusion

The results of the blinded listening experiment show that spatial AI, a new method for isolating target voices in loud, multi-speaker environments, seems to offer significant advantages over today's state of the art technology used throughout the industry. The investigational device was preferred by 95% of participants over all 5 top-spec competitors. Future research should explore the degree to which these large, consistent in-lab preferences translate to real world outcomes for hearing aid wearers. It is our recommendation that clinicians can and should prioritize comparative listening in noisy scenarios into the hearing aid selection process; in a world where all manufacturers are marketing their latest innovations, experiments like the one done here will help both clinicians and consumers find the signal in the noise. Returning to comparative trials as part of the hearing aid fitting/selection process will bring us back to the future thanks to the introduction of AI.

References:

American Speech-Language-Hearing Association. (2021). Attitudes and actions towards hearing health: Summary report of U.S. adults ages 18+. https://www.asha.org/siteassets/bhsm/2021/asha-bhsm-2021-report.pdf

Bentler RA. Effectiveness of directional microphones and noise reduction schemes in hearing aids: a systematic review of the evidence. J Am Acad Audiol. 2005 Jul-Aug;16(7):473-84. doi: 10.3766/jaaa.16.7.7. PMID: 16295234.

Bisgaard N, Vlaming MS, Dahlquist M. Standard audiograms for the IEC 60118-15 measurement procedure. Trends Amplif. 2010 Jun;14(2):113-20. doi: 10.1177/1084713810379609. PMID: 20724358; PMCID: PMC4111352.

Chung JS, Nagrani A, Zisserman A, VoxCeleb2: Deep Speaker Recognition INTERSPEECH, 2018.

Buchholz JM, Weisser A. (2019, January 13). Ambisonics Recordings of Typical Environments (ARTE) Database (1.0.0). Zenodo. https://doi.org/10.5281/zenodo.2261633

Cox RM, Johnson JA, Xu J. Impact of Hearing Aid Technology on Outcomes in Daily Life I: The Patients' Perspective. Ear Hear. 2016 Jul-Aug;37(4):e224-37. doi: 10.1097/AUD.0000000000000277. PMID: 26881981; PMCID: PMC4925253.

Dong R, Liu P, Tian X, Wang Y, Chen Y, Zhang J, Yang L, Zhao S, Guan J, Wang S. Influences of noise reduction on speech intelligibility, listening effort, and sound quality among adults with severe to profound hearing loss. Front Neurosci. 2024 Jul 23;18:1407775. doi: 10.3389/fnins.2024.1407775. PMID: 39108313; PMCID: PMC11301946.

Johnson, J., Xu, J., & Cox, R. (2016). Impact of hearing aid technology on outcomes in Daily Life II: Speech understanding and listening effort. Ear and Hearing, 37 (5), 529-540. https://doi.org/10.1097/AUD.0000000000000327

Morris C, Lovchinsky I, Callahan C, Agmon N, Lyngaas K, Malkin I, Shapiro W, Svirsky M. Spatial AI Improves Speech Intelligibility for Hearing Aid Wearers in Challenging Multi-talker Noise. Currently in peer review

Picou EM. Hearing Aid Benefit and Satisfaction Results from the MarkeTrak 2022 Survey: Importance of Features and Hearing Care Professionals. Semin Hear. 2022 Dec 1;43(4):301-316. doi: 10.1055/s-0042-1758375. PMID: 36466566; PMCID: PMC9715311.

Picou EM, Moore TM, Ricketts TA. The Effects of Directional Processing on Objective and Subjective Listening Effort. J Speech Lang Hear Res. 2017 Jan 1;60(1):199-211. doi: 10.1044/2016_JSLHR-H-15-0416. PMID: 28114609.

Valderrama, J.T., Mejia, J., Wong, A. et al. Reduced listening effort with adaptive binaural beamforming in realistic noisy environments. Sci Rep 15, 17998 (2025). https://doi.org/10.1038/s41598-025-95045-3

Ventry IM, Weinstein BE. The hearing handicap inventory for the elderly: a new tool. Ear Hear. 1982 May-Jun;3(3):128-34. doi: 10.1097/00003446-198205000-00006. PMID: 7095321.

Wu YH, Stangl E, Chipara O, Hasan SS, DeVries S, Oleson J. Efficacy and Effectiveness of Advanced Hearing Aid Directional and Noise Reduction Technologies for Older Adults With Mild to Moderate Hearing Loss. Ear Hear. 2019 Jul/Aug;40(4):805-822. doi: 10.1097/AUD.0000000000000672. PMID: 30379683; PMCID: PMC6491270.