Phonak Field Study News.

Spheric Speech Clarity significantly improves speech understanding in a multi-talker scenario and reduces spatial listening effort

This study, conducted in Shanghai, China compared two advanced noise reduction solutions available in Phonak Audéo Sphere Infinio and found that Spheric Speech Clarity surpassed StereoZoom 2.0, delivering up to 31% better speech intelligibility and a 35% reduction in listening effort, particularly for lateral speech sources.

Tian, X., Guan, J., Latzel, M., Kuehnel, V. October 2025

Introduction

Individuals with hearing loss face significant challenges in everyday communication, especially in environments with high background noise. The ability to recognize speech in such settings is crucial for maintaining social interactions, work productivity, and quality of life (Healy & Yoho, 2016). In noisy environments, the presence of multiple competing sound sources, such as background chatter, traffic noise or music can mask important speech cues, making it difficult for hearing-impaired individuals to distinguish speech from noise (Hawley et al., 2004; Auerbach & Gritton., 2022).

Speech recognition in noise is not only a function of overall signal-to-noise ratio (SNR) but also of the listener's ability to localize sound sources and attend to desired signal while ignoring others. This becomes problematic when the speech signal originates from an unexpected direction, leading to reduced intelligibility. For individuals with hearing loss, the difficulty in processing speech from multiple directions can limit their ability to participate in conversations, particularly in noisy situations such as restaurants, parties, or family gatherings (Shinn-Cunningham & Best, 2008).

Hearing aids have long utilized directional microphones to improve speech intelligibility in noise. Conventional directional microphones or monaural beamformers improve SNR by focusing on sounds from the front, where the primary speech source is often located (Bentler, 2005). More

advanced approaches, such as binaural beamformers, extend this concept by using two microphones on each hearing aid to create a four-microphone network via a wireless link. This configuration allows for a narrower beam directed to the front, further improving SNR compared to a monaural beamformer (Derleth et al., 2021). However, beamformers are less effective when speech comes from non-front angles, and this is especially true for binaural beamformers (Best, 2015).

Deep neural networks (DNNs) have rapidly gained attention as a promising approach to noise reduction in hearing aids. Traditional methods, such as spectral subtraction or Wiener filtering, rely on predefined rules and assumptions about noise characteristics. These approaches often struggle in environments where noise varies unpredictably. In contrast, DNNs are trained on diverse datasets of real-world speech and noise scenarios. This data-driven approach allows DNNs to suppress noise while preserving subtle speech features, such as tonal quality and temporal patterns, which are often degraded by traditional methods (Zheng et al., 2023).

The recent introduction of the Phonak Audéo Sphere Infinio hearing aids marks a significant milestone in applying DNN technology in the hearing aid industry. These devices utilize a dual-processing approach, integrating two distinct chips: - a dedicated Artificial Intelligence (AI) chip known as DEEPSONIC and a conventional signal processing chip, ERA. While ERA manages conventional noise reduction algorithms (e.g., Dynamic Noise Cancellation) and other hearing aid functions, the DEEPSONIC chip was designed with the [1]specific architecture[1] needed to run Spheric Speech Clarity (SSC), a DNN algorithm for real-time sound processing.

Technical measurements have shown that the effectiveness of SSC is not dependent upon the location of the target speech, unlike traditional beamformers that rely on a front-facing directional response (Raufer et al., 2024). This characteristic suggests SSC may surpass binaural beamformers, such as StereoZoom 2.0 (SZ 2.0) in complex listening situations. This study aims to compare the effectiveness of SSC to SZ 2.0 in Phonak Audéo Sphere Infinio hearing aids. Specifically, it evaluates their impact on speech recognition and listening effort when speech is presented from both the front and side. Study results will provide insight as to whether SSC offers significant advantages in multi-talker noisy environments where SZ 2.0 faces limitations.

[1] This architecture is covered by US Patent No. 12108219 B1.

Methodology

Participants

Seventeen participants (11 males and 6 females) aged 46 to 80 years (mean age: 68.2 years, SD = 10.5) with moderate to severe sensorineural hearing loss were recruited for the study. Hearing thresholds were measured using pure-tone audiometry, with average thresholds ranging from 40 to 80 dB HL across frequencies of 500 to 4000 Hz (Figure 1). All participants were native speakers of Mandarin.

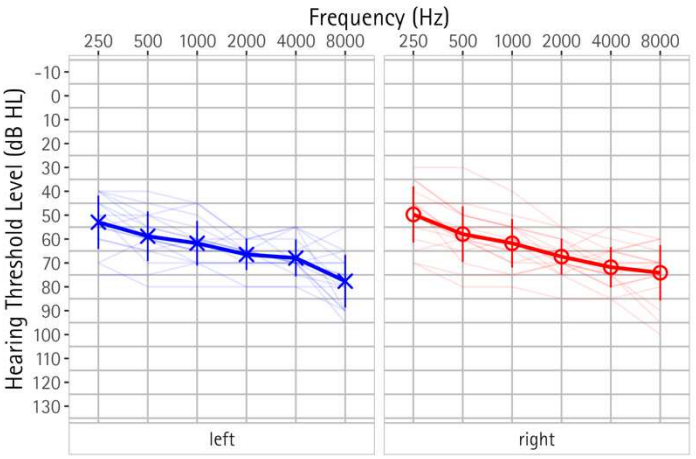

Figure 1. Mean hearing thresholds +/- one standard deviation for the left (blue) and right (red) ears.

Hearing aid fitting

Participants were fitted with Phonak Audéo Sphere Infinio 90 hearing aids, equipped with M or P receiver based on Target's recommendations. Custom SlimTips with Target-prescribed vent size were used to ensure appropriate acoustic coupling. The devices were programmed to Adaptive Phonak Digital Contrast 3.0 at 100% gain. A feedback test was performed to confirm that the acoustic coupling was well-sealed and there were no gain limitations. SoundRecover 2 (i.e., adaptive frequency compression) was disabled for the study. Two manual programs were created for comparison:

P1. Spheric Speech in Loud Noise (SSiLN): Fixed Directional [12]; the SSC feature was set to the default level, with all participants using a strength of 5, and other features remained at their default settings.

P2. Speech in Loud Noise (SiLN): SZ 2.0 [24]; other features remained at their default settings.

Laboratory Setup

The study took place at Sonova Audiology Research Center - Shanghai (ARC-SH). The test environment consisted of 12 loudspeakers arranged in a circle, positioned at equal intervals of 30° around the listener. The loudspeakers were mounted at ear level and placed at a distance of 1.4 m from the listener (Figure 2).

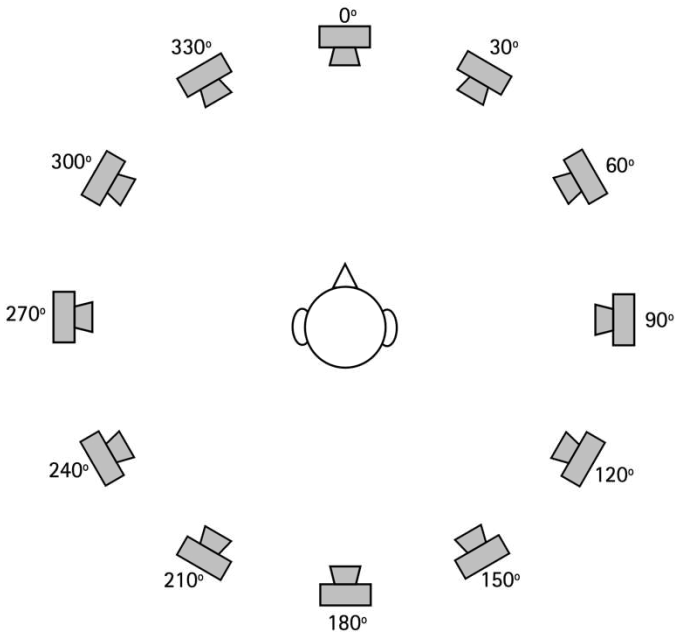

Figure 2. 12 Genelec 8030 loudspeakers were evenly placed in a circle with a 1.4 m radius. Participants sat in the center of the circle, facing a touch screen positioned in front of them.

Speech Intelligibility

Participants were seated at the center of a loudspeaker array with a flat-mounted touch screen placed on the table in front of them. The screen displayed 40 response options arranged in a four-row by ten-column matrix. Each button was labeled with a combination of one of the four colors (blue, green, red, or yellow) and one of the ten digits (1-10) from the Chinese Coordinate Response Measure (C-CRM; Wang et al., 2019) response set. The C-CRM speech material consists of sentences structured as "Ready, [call sign] [color] [number]" in Mandarin. The call sign identifies the target sentence, and the color-number pair represents the response.

Participants began the session by completing a speech reception threshold (SRT) test. The test setup included cafeteria noise presented from all 12 loudspeakers at an overall level of 70 dB(A), with C-CRM sentences voiced by a female talker originating from 0° azimuth (Figure 2). The one-down, one-up tracking algorithm was employed to estimate the point on the psychometric function corresponding with 50% correct identification (Levitt, 1971). Next, participants completed a muti-talker speech recognition in noise task adapted from Heeren et al. (2022). The task used the same noise setup as in the SRT test along with three distinct voices: a female talker at 0°, a male talker at 60°, and another male talker at 300°. Noise was again presented at an overall level of 70 dB(A), and speech was presented at 3 dB above the participants' SRT. Sentences with the call sign "Wang Li" served as target sentences. Participants selected the color-number pair spoken by the last talker who said "Wang Li".

The task required participants to simultaneously attend to the color-number pair of the current target sentence and the call sign of the following sentence. To increase task complexity, a 0.6-second overlap between adjacent sentences ensured that the color-number pair overlapped with the beginning of the following sentences from a different talker. This design limited the number of active talkers to two at any given time (Figure 3). Each test block included 21 target sentences (7 per talker), with each target sentence followed by two or three distractor sentences. Participants completed the task twice, once during the initial test and again for a retest conducted several days later.

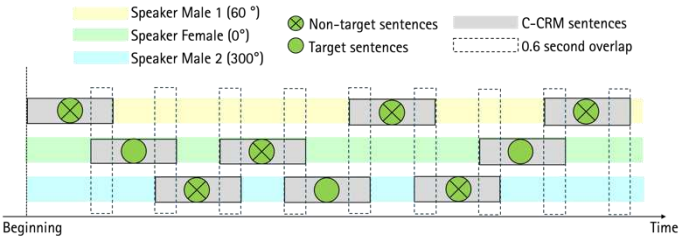

Figure 3. Diagram of the multi-talker speech recognition test procedure including three speakers positioned at 60° (male 1), 0° (female 1), and 300° (male 2). Dashed vertical lines indicate overlapping duration between adjacent talkers' speech intervals.

Listening Effort

The listening effort test was conducted using Adaptive Categorical Listening Effort Scaling (ACALES; Krueger et al., 2017) method. The test featured two female Mandarin Chinese matrix sentences (Hu et al., 2018) presented randomly from 0°, 60°, and 300°, accompanied by cafeteria noise from all directions. Participant rated the effort required to follow the speaker at the current SNR using a 14-point categorical rating scale from "effortless" to "extreme effort," with an additional "only noise" option to be used if the participant could not detect the presence of target speech. The loudspeaker setup mirrored that of the speech recognition in noise test. The initial SNR was set to 5 dB with the overall noise level at 70 dB(A).

Results

Speech intelligibility

Speech intelligibility was calculated as a percent correct for each individual target location, as well as an average across all locations (Figure 4).

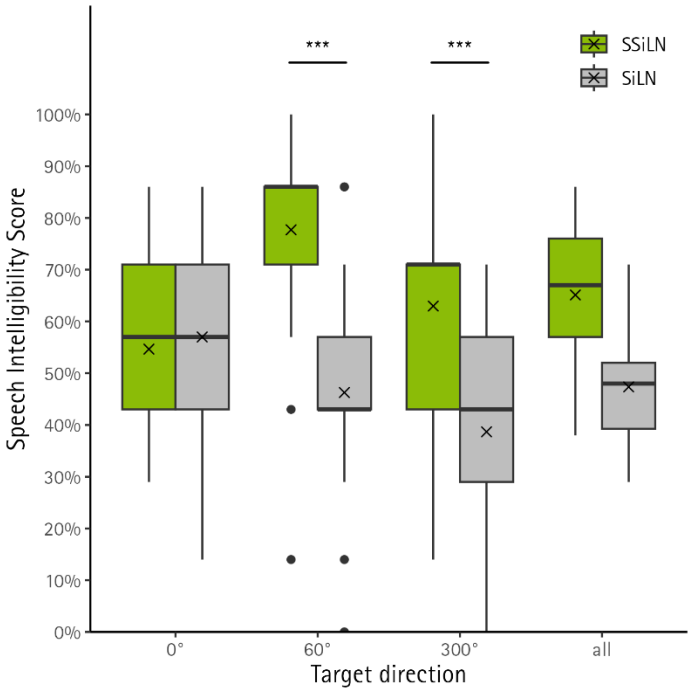

Figure 4. Speech intelligibility score (%) across different target directions. Mean score is represented by a cross.

The data were statistically analyzed in the R software (version: 4.4.1; R Core Team 2021) using generalized linear mixed effect model (GLMM) with binomial family and logit link function via the *lme4* package. The model included hearing aid program, target direction, and test sequence, along with their interaction, as fixed effects. Participant was included as random intercept to account for individual variability. The analysis of variance (ANOVA) results revealed significant interactions between hearing aid program and target direction ($F(2, 196) = 17.05$, $p < .001$), indicating the effect of hearing aid program on speech intelligibility varies depending on the direction of target speaker.

Target direction	Hearing aid program	Mean speech intelligibility	95% Confidence Interval (%)	P value
0°	SiLN	57.1%	[41.1, 74.4]	1.000
0°	SSiLN	54.6%	[36.3, 71.9]	
60°	SiLN	46.2%	[28.7, 62.1]	< .001
60°	SSiLN	77.9%	[63.5, 88.1]	
300°	SiLN	38.6%	[23.4, 57.7]	< .001
300°	SSiLN	63.1%	[46, 79.4]	

Table 1. Speech intelligibility score across conditions

Table 1 presents the predicted probabilities for speech intelligibility under different hearing aid programs at three target directions. The SSiLN condition significantly improved intelligibility by 31% at 60° and 24% at 300°, compared to the SiLN condition ($p < .001$).

Listening effort

Subjective listening effort data were analyzed using a linear mixed-effects model (LMM). The model included fixed effects of hearing aid program, target direction, categorial effort rating, test sequence, and their interaction, while participants served as random effects. The model revealed a significant main effect of hearing aid program ($\beta = 2.20$, $p < .001$) and significant interaction effects between hearing aid program and target direction for speaker at 60° ($\beta = 2.00$, $p < .001$) and speaker at 300° ($\beta = 1.92$, $p < .001$).

Post-hoc comparisons indicated that the SSiLN leads to 2.2 dB, 4.2 dB, and 4.1 dB lower SNRs than the SiLN for speaker at 0° (SE = .243, $t = 9.03$, $p < .001$), 60° (SE = .247, $t = 16.99$, $p < .001$), and 300° (SE = .238, $t = 17.3$, $p < .001$), respectively. These lower SNRs suggest that participants could tolerate less favorable SNR conditions without experiencing increased levels of listening effort, reflecting a reduction in cognitive demands during listening when using SSiLN. Figure 5 illustrates this relationship by showing a decline in SNR with increasing listening effort, where SSiLN consistently performs better than SiLN, particularly at higher effort levels.

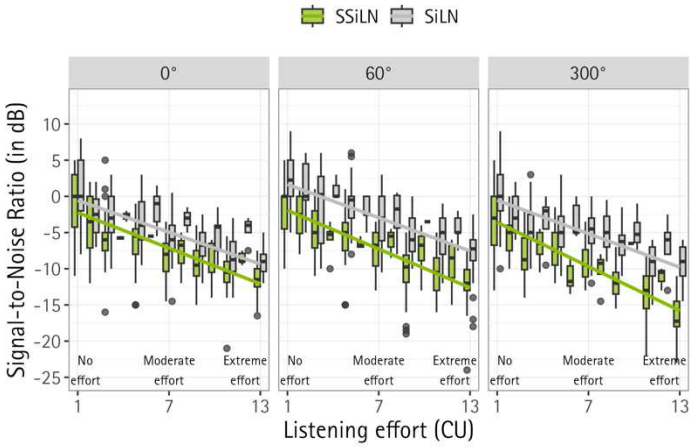

Figure 5. Signal-to-noise ratio (SNR) as a function of listening effort (CU) for SSiLN (green) and SiLN (grey) across three target directions (0°, 60°, and 300°).

The reduction of listening effort in SNR values can be calculated into percentage changes by assuming a linear relationship between categorical units and perceived listening effort, where each unit contributes equally. Under this assumption, the observed SNR reductions of 2.2, 4.2, and 4.1 dB correspond to 2.4, 4.6 and 4.5 categorical unit changes, respectively. These reductions, in turn, translate to percentage improvements in listening effort of approximately 18%, 35%, and 34%, respectively.

Discussion and conclusion

The current study demonstrates the remarkable ability of the new Spheric Speech in Loud Noise program to preserve speech cues from off-center sound sources (60° and 300°). These outcomes align with the technical measurements showing that SSC alone provided 5.8 to 6.9 dB SNR improvement for speech originating from angles other than

0° (Raufer et al., 2024). In this study, no significant difference was observed for speech intelligibility at 0°, suggesting that SSC and SZ 2.0 can create similar levels of benefit for on-axis speech. As speech moves off axis, SSC uniquely maintains this level of benefit, while the benefit for SZ 2.0 is limited (as expected).

While speech understanding for both programs was comparable for speech from 0°, listening effort data slightly favored SSC. This emphasizes the distinction between speech comprehension and the cognitive effort required to achieve it.

A unique aspect of the study was the implementation of the ACALES test with spatial variability in the presentation of target speech stimuli. This design reflects real word listening environments where speech sources are rarely fixed to a single frontal location. The results suggest that SSC is more robust in handling spatial variability and offering clear advantages in realistic communication scenarios involving multiple or off-center sound sources.

References

Auerbach, B. D., & Gritton, H. J. (2022). Hearing in complex environments: auditory gain control, attention, and hearing loss. Frontiers in neuroscience, 16, 799787.

Bentler, R. A. (2005). Effectiveness of directional microphones and noise reduction schemes in hearing aids: A systematic review of the evidence. Journal of the American Academy of Audiology, 16(07), 473-484.

Best, V., Mejia, J., Freeston, K., Van Hoesel, R. J., & Dillon, H. (2015). An evaluation of the performance of two binaural beamformers in complex and dynamic multitalker environments. International Journal of Audiology, 54(10), 727-735.

Derleth, P., Georganti, E., Latzel, M., Courtois, G., Hofbauer, M., Raether, J., & Kuehnel, V. (2021, August). Binaural signal processing in hearing aids. In Seminars in Hearing (Vol. 42, No. 03, pp. 206-223). Thieme Medical Publishers, Inc..

Hawley, M. L., Litovsky, R. Y., & Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. The Journal of the Acoustical Society of America, 115(2), 833-843.

Healy, E. W., & Yoho, S. E. (2016, August). Difficulty understanding speech in noise by the hearing impaired: underlying causes and technological solutions. In 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (pp. 89-92). IEEE.

Heeren, J., Nuesse, T., Latzel, M., Holube, I., Hohmann, V., Wagener, K. C., & Schulte, M. (2022). The Concurrent OLSA test: A method for speech recognition in multi-talker situations at fixed SNR. Trends in Hearing, 26, 23312165221108257.

Hu, H., Xi, X., Wong, L. L., Hochmuth, S., Warzybok, A., & Kollmeier, B. (2018). Construction and evaluation of the Mandarin Chinese matrix (CMNmatrix) sentence test for the assessment of speech recognition in noise. International Journal of Audiology, 57(11), 838-850.

Krueger, M., Schulte, M., Brand, T., & Holube, I. (2017). Development of an adaptive scaling method for subjective listening effort. The Journal of the Acoustical Society of America, 141(6), 4680-4693.

Levitt, H. C. C. H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical society of America*, *49*(2B), 467-477.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. Language and Cognitive processes, 27(7-8), 953-978.

Raufer, S., Kohlhauer, P., Jehle, F., Kühnel, V., Preuss, M., Hobi, S. (2024). Spheric Speech Clarity proven to outperform three key competitors for clear speech in noise. Phonak Field Study News retrieved from https://www.phonak.com/evidence

R Core Team. (2021). A language and environment for statistical computing

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. Trends in amplification, 12(4), 283-299.

Wang, Y., Lu, Z., Yang, X., & Liu, C. (2019). Measuring Mandarin speech recognition thresholds using the method of adaptive tracking. Journal of Speech, Language, and Hearing Research, 62(6), 2009-2017.

Zheng, C., Zhang, H., Liu, W., Luo, X., Li, A., Li, X., & Moore, B. C. (2023). Sixty years of frequency-domain monaural speech enhancement: From traditional to deep learning methods. Trends in Hearing, 27, 23312165231209913.

Acknowledgements

The authors would like to thank Kevin Seitz-Paquette for his valuable contribution in editing and refining the content of this document.

Authors and investigators

Xin Tian, Clinical Audiology Manager

Xin Tian joined the Sonova Audiology Research Center in Shanghai as a Clinical Audiology Engineer in 2015. With a background in audiology, he has worked on various research and development projects related to hearing aid technologies, collaborating closely with both HQ and local teams. His expertise spans product testing, clinical evaluation, and optimizing user experience.

Jingjing Guan, PhD., Senior Director ARC Shanghai

Jingjing holds a Ph.D. degree from University of Texas at Austin and worked in Texas Tech University Health Sciences Center as an assistant professor and a licensed audiologist. Jingjing started her journey in Sonova Audiology Research Center in Shanghai in 2018. Her main research interests include psychoacoustics and hearing aid outcome measures.

Matthias Latzel, Ph.D., Senior Expert Clinical Studies

Dr. Matthias Latzel studied electrical engineering in Bochum and Vienna in 1995. After completing his Ph.D. in 2001, he carried out his PostDoc from 2002 to 2004 in the Department of Audiology at Giessen University. He was the head of the Audiology department at Phonak Germany from 2011. From 2012 to 2022 he has been working as the Clinical Research Manager for Phonak AG, Switzerland. Since 2022 he is senior expert clinical studies for Sonova AG.

Volker Kühnel, PhD., Principal Expert Hearing Performance

Volker Kühnel, PhD, holds a doctorate in physics and completed his studies in 1995. From 1995 to 1997 he worked in Oldenburg as a research assistant in the Medical Physics group of Prof. B. Kollmeier. Since 1998, he has been working at Phonak/Sonova in product development on the audiological design at the interface between hearing aid algorithms and fitting software. His work focuses on the audiological quality of hearing systems to achieve maximum customer benefit.

Phonak Field Study News.

One-page summary

Spheric Speech Clarity significantly improves speech understanding in a multi-talker scenario and reduces spatial listening effort

This study found that Spheric Speech Clarity surpassed StereoZoom 2.0, delivering up to 31% better speech intelligibility and a 35% reduction in listening effort, particularly for lateral speech sources.

Tian, X., Guan, J., Latzel, M., Kuehnel, V. October 2025

Key highlights

- Spheric Speech Clarity (SSC) provided up to 31% better speech intelligibility compared to StereoZoom 2.0 (SZ 2.0) when speech came from lateral locations (60° and 300°).
- Participants with SSC experienced up to 35% less listening effort compared to SZ 2.0. This reduction was especially noticeable when speech came from the sides.
- APD 3.0 offers new frequency response curves designed to enhance general sound quality in all programs.

Considerations for practice

- SSC can be beneficial for users in challenging listening environments, such as noisy gatherings or crowded public spaces. HCPs might consider prioritizing SSC for clients who struggle in multi-talker situations or other complex listening environments, like in busy cafés.
- The reduction in listening effort with SSC implies that patients may experience less cognitive strain during conversations. This could be useful for individuals who report mental fatigue or difficulty following speech in complex environments.