Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Speech intelligibility improvement effect of a soundbar equipped with a hearing-aid algorithm

DEIDDA SURIAUT Gianni
Seoul Audiology Education Center
Chae-yeon Yu
Seoul Audiology Education Center
Moo Kyun Park
aseptic@snu.ac.kr

Seoul National University College of Medicine

Article

Keywords: Presbycusis, Soundbar, Speech intelligibility, Hearing assistance, Sentence recognition, Older adults

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11232286/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. This work was supported by a research grant from JD Solution, the manufacturer of the soundbar evaluated in this study, which also provided the device on loan. JD Solution had no role in the study design, data collection, data analysis, interpretation of the data, or preparation of the manuscript. The authors declare no other financial or non-financial competing interests.

Speech intelligibility improvement effect of a soundbar equipped with a hearing-aid algorithm.

DEIDDA SURIAUT Gianni[1,3], Chae-yeon Yu[1,4], Moo Kyun Park[2,4,*]

1. Seoul National University College of Medicine, Seoul Audiology Education Center, Seoul, Republic of Korea.

2. Seoul National University College of Medicine, Department of Otorhinolaryngology-Head and Neck Surgery, Seoul, Republic of Korea.

3. Audiocampus, Université de Montpellier, Montpellier, France

4. Interdisciplinary Program in Neuroscience, College of Natural Sciences, Seoul National University, Seoul, Republic of Korea.

*Correspondence: Moo Kyun Park; email: aseptic@snu.ac.kr

Abstract

Age-related hearing loss can impair television speech understanding. We evaluated a soundbar incorporating an NAL-NL2-based hearing-assistance algorithm in 32 older adults (mean age, 69.6 years). Word and sentence recognition were measured in standard (M0) and hearing-assistance modes (M2 and M3) in quiet and noise, in a sound-treated booth and a simulated home-listening environment. Compared with M0, the assisted modes improved recognition, with the greatest gains for sentences in noise (M2: 16.4 percentage points; M3: 22.3 percentage points; both $p < 0.001$). Principal component analysis and k-means clustering identified a more severely hearing-impaired subgroup (n = 8), which showed larger sentence-recognition gains in noise (up to 65.3 points) than a subgroup with better hearing (n = 24; up to 8.0 points). A mixed-effects model detected a mode-by-cluster interaction ($p < 0.001$), and greater auditory impairment was associated with larger gains in continuous regression analyses ($R^2 = 0.40–0.63$; all $p < 0.001$). Subjective ratings generally favored the assisted modes. These findings suggest that the soundbar may improve speech intelligibility for older adults with hearing loss, particularly in noise.

Keywords

Presbycusis; Soundbar; Speech intelligibility; Hearing assistance; Sentence recognition; Older adults

I) Introduction:

Presbycusis, or age-related hearing loss, is one of the most widespread sensory impairments among older adults worldwide [1]. With the continued aging of the global population, the number of people affected by this condition is expected to increase considerably in the coming decades. Presbycusis is therefore becoming a growing public health concern. The Republic of Korea (South Korea) is a particularly striking case. In 2025, people aged 65 and over account for 20.3% of the country's total population, officially placing it among "super-aged societies." This proportion is expected to reach 30% by 2036 and 40% by 2050 [2]. This rapid demographic shift is expected to further increase the burden of presbycusis within the country. This decline in hearing can lead to numerous difficulties for older adults, including difficulty understanding speech while watching television [3, 4]. It is important to note that in South Korea, nearly 66% of people with hearing loss do not wear hearing aids, as the use of a hearing aid remains heavily stigmatized in society [5]. It is therefore necessary to explore strategies and develop innovative devices to make daily life easier for these hearing-impaired individuals, while maintaining adequate stimulation of their auditory system [6].

In recent years, some manufacturers of high-power directional loudspeaker systems have begun adapting their expertise for consumer use. These systems were originally designed for security, emergency announcements, or military applications. Today, some manufacturers are developing soundbars that incorporate hearing-aid algorithms. This type of innovation is primarily intended for hearing-impaired seniors, to assist them while watching television. However, its use extends well beyond this single population. Indeed, a soundbar considerably improves the sound quality of the television for all members of the household, whether they have hearing loss.

This study evaluated speech intelligibility with a hearing-assistance soundbar in a sound-treated booth and a simulated home-listening setting, in quiet and in noise. We assessed recognition across amplification modes in the entire sample and in auditory profiles identified using principal component analysis (PCA) and k-means clustering. A mixed-effects model and regression on a continuous auditory-severity index were used to examine whether hearing status was associated with amplification benefit. We also assessed participants' ratings of the listening modes.

II) Materials and Methods:

a) Soundbar:

We used the soundbar developed by the Korean company JD Solution. The "Haru'nje" soundbar was first unveiled at the KOBA trade show in Seoul in 2025 (International Broadcast, Audio & Lighting Equipment Exhibition). It also received the Best of Audio Innovation Award 2025 at the IFA trade show in Berlin (the world's largest trade show dedicated to consumer electronics and home technology), which gave it a degree of international visibility. This soundbar was not primarily designed for home theater use, but rather to help older adults and/or people with hearing loss better understand television dialogue.

To achieve this, the company's engineers worked on two innovative technologies. First, a directional audio technology that focuses sound delivery precisely toward the listener. By directing sound in this rigorous way, intelligibility is improved without needing to increase the volume. This technology cannot be controlled by the user; it is built into the design of the soundbar. Second, they incorporated a hearing-aid algorithm based on the NAL-NL2 prescription method (National Acoustic Laboratories, Non-Linear version 2) into the soundbar [7, 8]. The "Haru'nje" offers several listening modes with different degrees of amplification that can be selected by the user according to their needs and preferences. There are five different modes: "Music," "News," "M1," "M2," and "M3." The "Music" mode is the standard mode, with uniform sound amplification. In this study, we refer to this mode as "M0." The "News" mode slightly amplifies frequencies from 600 Hz to 10 kHz while slightly reducing low frequencies. We did not use the "News" mode in this study. Finally, the "M1," "M2," and "M3" modes are intended for people with hearing loss. They provide substantial amplification across the 0.6-10 kHz range, along with a reduction in low frequencies. Moving up to a higher assistance level results in progressively greater gain applied to the targeted frequencies and progressively lower gain on low frequencies (>0.5 kHz). The manufacturer indicates that this reduction in low-frequency sound, which varies in magnitude and is not native to the NAL-NL2 formula, was designed to maintain a consistent overall sound level across modes. This way, seniors (or people with hearing loss), who tend to raise the television's overall volume, can instead select

a higher hearing-assistance level. This approach aims to improve speech intelligibility while limiting the increase in perceived sound level for other people in the room (particularly those with normal hearing). In these three modes, a dynamic compression system is also introduced, applying greater gain to soft sounds than to loud ones. This compression enhances the perception of consonants, which carry a significant portion of the phonemic information needed for speech comprehension. The sound levels used in our study remained below the compression's activation threshold. As a result, the amplification applied was linear across all measurements taken.

The manufacturer-provided frequency responses are shown in Supplementary Fig. S1.

b) Participants:

Recruitment of participants was carried out with the help of Seoul National University Hospital. This hospital has an office where patients can independently register on lists to take part in scientific research studies. We used these lists to contact participants. However, the lists do not contain any audiometric data. We had no way of knowing the hearing status of a potential participant in advance. To maximize our chances of recruiting participants with hearing loss, we exclusively selected participants aged 65 or older. All participants had to be native Korean speakers. We also excluded any participant suspected of having cognitive impairment. All participants were compensated for taking part in the study.

For this study, 32 subjects were recruited; the mean age was 69.6 ± 3.1 years (95% CI).

c) Test Procedure:

The study was approved by the Institutional Review Board of Seoul National University Hospital (approval no. 2506-019-1648). Written informed consent was obtained from all participants before study procedures. All procedures were conducted in accordance with the Declaration of Helsinki and applicable institutional guidelines.

Participants completed a wide range of tests divided into three phases. The order in which tests were administered within each phase was randomized for each participant to limit order and fatigue effects. In addition, participants had no way of knowing which amplification mode was being used during each test. All speech tests were administered by the same examiner. Whenever speech had to be presented through the audiometer's microphone, a female voice was used. In addition, the M1 mode was not evaluated in this study. Adding this mode alongside M2 and M3 would have been possible but would have lengthened test duration and increased participant fatigue. We therefore chose to focus measurements on the M2 and M3 modes, considered a priori more representative of amplification intended for hearing-impaired profiles.

The first phase took place inside a soundproof booth. In this booth, the ambient sound level of 40 dB SPL was measured using a Cirrus CR:152B sound level meter. We began with the tests common to all audiology studies at Seoul National University: the screening tests. We used an AC40 audiometer along with DD45 headphones. The screening tests consisted of pure-tone air-conduction audiometry. For speech recognition tests, we

collected the WRS at MCL and the SRT. Pure-tone air-conduction audiometry determined the hearing thresholds for each of the subject's ears at the following frequencies: 125 Hz, 250 Hz, 500 Hz, 750 Hz, 1000 Hz, 1500 Hz, 2000 Hz, 3000 Hz, 4000 Hz, 6000 Hz, and 8000 Hz. The stimuli used were warble pure tones. The Word Recognition Score (WRS) is a speech audiometry test using monosyllabic words delivered through the audiometer (pre-recorded lists), presenting 25 words per list [9]. Words were presented at the participant's Most Comfortable Level (MCL). To determine this level, the examiner spoke to the participant through the audiometer's microphone and adjusted the sound level according to the participant's feedback, until reaching the level judged most comfortable. Presenting word lists at MCL allows estimation of the participant's maximum speech recognition performance and can also provide information on certain cognitive functions involved in speech processing [10]. The Speech Recognition Threshold (SRT) was determined using bisyllabic words presented orally by the examiner through the audiometer's microphone. The presentation intensity was progressively adjusted to determine the minimum sound level at which the participant correctly recognized 66% of the words presented. Next, we connected the audiometer to the soundbar so that it could deliver warble pure tones as well as speech. The soundbar was placed facing the participant at 1.20 m, at head height. We performed free-field pure-tone audiometry using modes M0, M2, and M3 at the following frequencies: 500 Hz, 1000 Hz, 2000 Hz, 3000 Hz, 4000 Hz, and 8000 Hz. In addition, we collected WRS scores as well as Sentence Recognition Scores (SRS), both presented at 65 dB HL for modes M0 and M2. SRS scores were determined from ten sentences presented orally by the examiner, representing a total of 40 words [11]. This first phase was conducted under so-called laboratory conditions. Indeed, the acoustic and sound-delivery conditions do not reflect those encountered in everyday life. The low ambient noise level (40 dB SPL), the low reverberation, and the optimized positioning of the soundbar create an environment far removed from those typically encountered in a home setting.

In the second phase, we sought to come as close as possible to the listening conditions encountered when watching television in a home environment. To do this, the tests were carried out in our usual office, a furnished room with acoustic characteristics closer to a domestic setting. The soundbar was placed at a distance of 2.5 m facing the participant and, this time, positioned below chest height. This allowed us to more faithfully reproduce the usual conditions of television listening. We measured WRS at 65 dB SPL using M0 and M2, as well as SRS at 65 dB SPL using M0, M2, and M3. The M3 mode was only tested for SRS. Indeed, sentence recognition is more representative of television listening than isolated word recognition. In addition, this decision helped limit test duration in order to reduce participant fatigue while still collecting as much data as possible. We measured WRS and SRS both in silence and in noise. The word lists consisted of 25 monosyllabic words [9], and the sentence lists comprised 10 sentences totalling 40 words [11]. Words and sentences were played from a phone connected via Bluetooth to the soundbar. The speech stimuli came from a recording made by a male voice and were presented at a level of 65 dB SPL, measured using a sound level meter at the position of the participant's head [9, 11]. The ambient sound level of 50 dB SPL for the silence condition was measured using the sound level meter. For the noise condition, we used a four-talker cocktail-party noise spoken in Korean. It was played at a level of 65 dB SPL, measured with the sound level meter at the position of the participant's head. To reproduce a more realistic sound environment, the background noise was played through two speakers placed at the rear left and rear right of the soundbar. The speakers were oriented toward the walls, promoting a predominantly reverberant sound field. These configurations were designed to approximate the listening conditions encountered when watching television in a quiet or noisy home environment.

The third and final phase consisted of collecting participants' subjective impressions using a questionnaire. This questionnaire included six items assessing perceived sound quality and listening satisfaction for each of the modes tested (M0, M2, and M3). The items were drawn from the International Outcome Inventory for Hearing Aids (IOI-HA) and the Abbreviated Profile of Hearing Aid Benefit (APHAB) [12, 13].

For each listening mode, participants were asked to give a score between 0 and 10 for the following items:

- Speech is easy to understand.
- The sound is comfortable to listen to.
- The sound quality seems natural.
- There is no reverberation causing sound overlap or reduced speech intelligibility.
- No sound causes discomfort or annoyance.
- I prefer this listening mode.

Higher scores assigned by the participant reflected a more favorable appreciation of the mode evaluated.

The two test settings are illustrated in Supplementary Fig. S2.

d) Statistical Analysis:

Screening test results were summarized separately for the better and worse ears to account for hearing asymmetry. The pure-tone average (PTA) for each ear was calculated as the mean threshold at 500, 1000, 2000, and 4000 Hz, according to the World Health Organization method [14]. Differences between ears were evaluated using Wilcoxon signed-rank tests.

First, we analysed the gain differences between the mean free-field audiogram thresholds according to the mode used. We then analysed the evolution of WRS and SRS as a function of amplification mode (M0, M2, M3) according to listening condition (silence at 40 dB SPL, silence at 50 dB SPL, noise at 65 dB SPL) using Wilcoxon signed-rank tests. Speech recognition tests (WRS and SRS) are initially expressed as percentages; however, we transformed them into RAU (Rationalized Arcsine Unit) values before any analysis [15]. This choice was made to correct the variance distortion caused by the ceiling effect observed in participants whose scores were already close to 100% with the unamplified mode (M0). These initial analyses were carried out across the entire sample in order to assess the overall influence of each amplification mode in our three listening conditions.

Second, we performed an unsupervised classification of participants based on their auditory profile. To do this, a principal component analysis (PCA) was carried out using four standardized variables (z-scores): the PTA of the better ear, the PTA of the worse ear, the WRS score in silence at 50 dB SPL in M0 mode, and the WRS in noise at 65 dB SPL in M0 mode. For speech recognition scores, we kept the RAU-transformed scores as variables in the PCA. Our choice to include both tonal variables (PTA) and speech recognition variables (WRS) is based on the observation that some participants may have good tonal hearing but poor word recognition

performance, particularly in the presence of noise [16]. WRS was preferred over SRS as the input variable for the PCA. Monosyllabic word recognition limits reliance on cognition and contextual anticipation, which can facilitate sentence recognition. WRS thus constitutes a more robust index of phonemic discrimination [17]. K-means clustering was then applied to the scores derived from the principal component analysis using Euclidean distance. The optimal number of clusters was determined using the elbow method, based on within-cluster inertia (WCSS) for a number of clusters ranging from 1 to 6. This method led to a two-cluster solution being selected, supported by the silhouette coefficient analysis, whose mean value was highest for two clusters (0.742). For each cluster, we analyzed the evolution of WRS (M0, M2) and SRS (M0, M2, M3) as a function of amplification mode, according to listening condition (silence at 40 dB SPL, silence at 50 dB SPL, noise at 65 dB SPL). Within-cluster Wilcoxon signed-rank tests were used to determine whether each amplification mode provided a significant improvement within each cluster according to listening condition.

Third, we sought to determine whether the effect of amplification differs between the two clusters. Since the clusters were formed based on hearing thresholds (PTA) and word recognition scores (WRS) in M0 mode, a simple comparison of raw SRS gains (post-score − pre-score) between clusters carries a potential bias: a cluster with lower initial performance generally has a greater margin for improvement, independently of the true effect of the amplification mode. Since the protocol included three repeated measures per participant (M0, M2, and M3), differences in evolution between clusters were therefore assessed using a linear mixed model, separately for silence and for noise. The model included the fixed effects of amplification mode, cluster, and their interaction (SRS RAU) ~ Mode × Cluster + (1 | Subject)). A significant interaction between amplification mode and cluster would indicate that the evolution of SRS scores differs according to cluster (and thus according to auditory profile). The normality of the model's residuals was checked using the Shapiro-Wilk test.

Fourth, we sought to build a predictive model of the benefit provided by the soundbar to move beyond a group-based classification. A multiple linear regression including the four variables derived from the PCA (better-ear PTA, worse-ear PTA, silence WRS at M0 (RAU), noise WRS at M0 (RAU)) was initially tested to assess their ability to predict individual gain according to listening condition. However, strong collinearity between these four predictors was revealed by the variance inflation factor (VIF = 3.1 to 4.8) and by an inter-variable correlation matrix (r = 0.65 to 0.84), making individual interpretation of the multivariate model's coefficients unstable. Rather than arbitrarily selecting one of these four correlated variables as the sole predictor, we chose to use the first principal component (PC1) from the PCA as a composite index of overall auditory profile severity. This approach was applied separately to the four possible combinations: silence gain and noise gain, for modes M2 and M3.

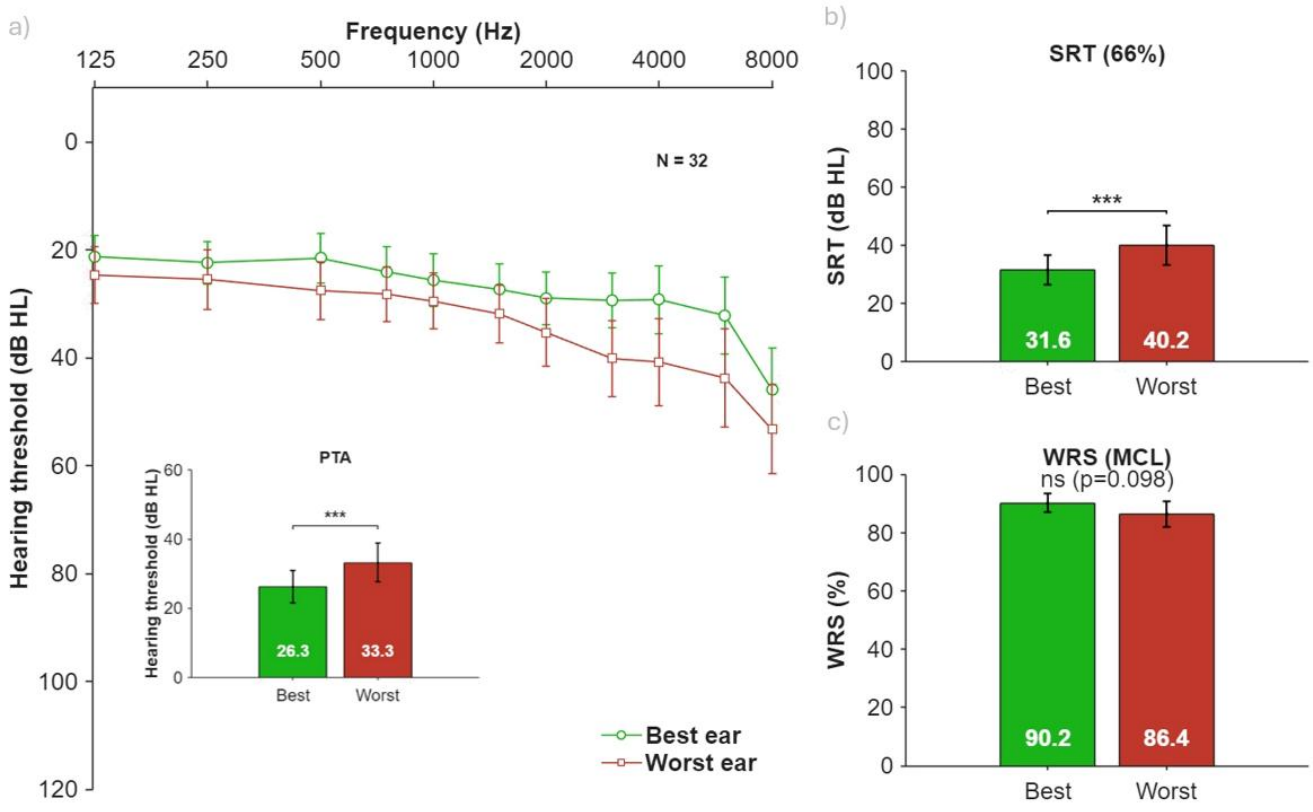
Fifth and finally, we sought to analyse the satisfaction questionnaire results item by item. This analysis was carried out both on the overall sample and separately within each of the two clusters identified by the PCA and k-means. We analysed whether perceived satisfaction with the amplification modes differed according to the participant's auditory profile.

The normality of the distributions of all the data was checked using the Shapiro-Wilk test. We noted that some sets of variables did not follow a normal distribution. For this reason, we decided to uniformly apply a Wilcoxon signed-rank test to all paired comparisons. This choice was made to maintain a consistent statistical methodology across the entire dataset. Statistical comparisons of SRS and WRS were carried out on scores expressed in RAU units. Although statistical comparisons were performed using non-parametric tests, we

retained a presentation as means ± 95% confidence intervals to facilitate graphical reading and comparison with the audiological literature. For the confidence intervals of the speech scores, we used a percentile bootstrap (resampling subjects with replacement, 2,000 replicates) rather than the classical Student's t approximation. This choice was made to prevent the interval bounds from exceeding the natural limits of percentages (0–100%). In addition, the figures present WRS and SRS scores as raw percentages, without the RAU transformation, also to facilitate graphical reading. For all comparisons involving simultaneous comparisons, a Bonferroni correction was applied. The relationship between the composite index PC1 and the speech recognition gain was assessed using simple linear regression. In the figures, significance levels are indicated as follows: ns = not significant; * : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$. Statistical analyses and figures were performed using MATLAB.

III) Results:

a) Baseline Audiological Characteristics of the Participants:

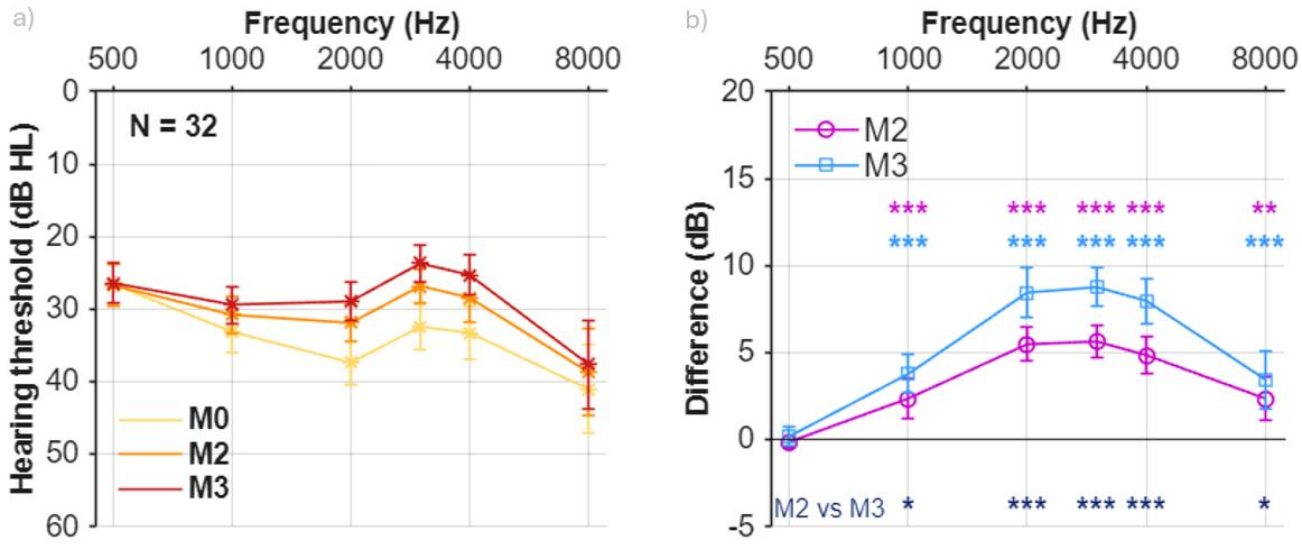


Baseline audiological characteristics are summarized in Figure 1. As shown in Figure 1a, the mean pure-tone audiogram demonstrates a bilateral high-frequency hearing loss. The mean PTA was 26.3 ± 4.6 dB HL for the better ear and 33.3 ± 5.6 dB HL for the worse ear, corresponding to a statistically significant interaural difference of 7 dB HL.

Similarly, the mean SRT differed by 8.6 dB HL between ears. The worse ear required a significantly higher presentation level (40.2 ± 6.7 dB HL) than the better ear (31.6 ± 5.0 dB HL) to achieve the speech recognition threshold.

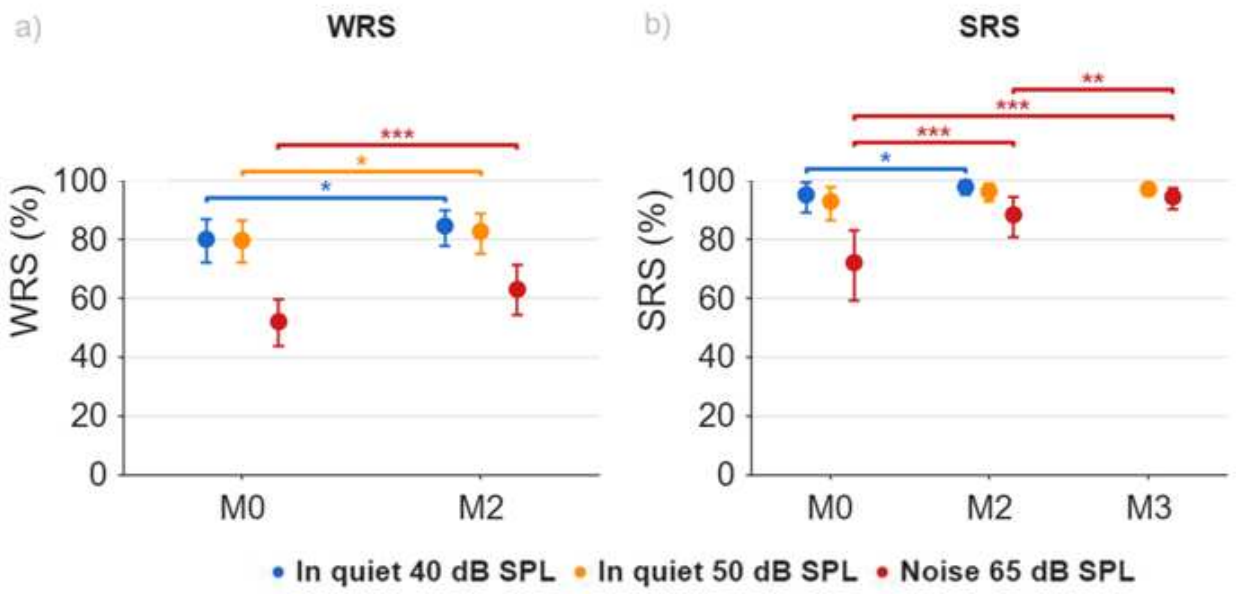
In contrast, when words were presented at each participant's Most Comfortable Level (MCL), word recognition performance was comparable between ears. The mean WRS was 90.2 ± 3.1% for the better ear and 86.4 ± 4.4% for the worse ear, with no statistically significant difference between them.

b) Effect of Amplification in the Overall Cohort:



As shown in Figure 2a, free-field pure-tone thresholds progressively improved with increasing levels of amplification. The mean PTA decreased from 32.6 ± 1.6 dB HL in the M0 mode to 29.5 ± 0.8 dB HL in M2, and further to 27.5 ± 0.7 dB HL in M3.

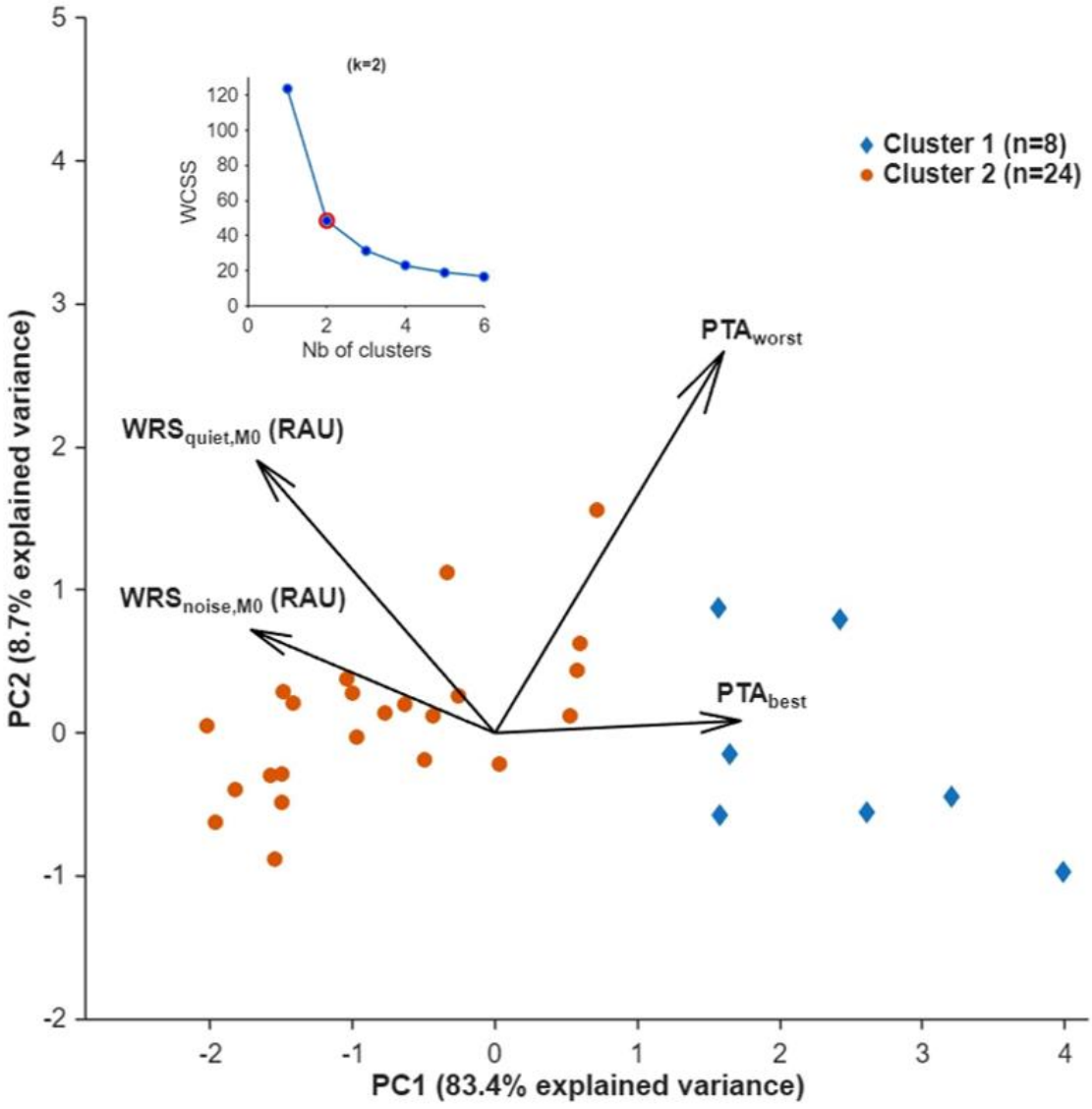
Figure 2b focuses on the comparison of gains achieved by the different modes as a function of frequency. Compared to M0, modes M2 and M3 provide a significant gain across the entire tested frequency spectrum, except at 500 Hz. This gain is more pronounced around the mid and high frequencies (1000, 2000, 3000, 4000, and 8000 Hz), with respective gains of 2.34, 5.46, 5.6, 4.84, and 2.34 dB for M2, and 3.75, 8.43, 8.75, 7.97, and 3.43 dB for M3. When comparing M2 with M3, the superiority of M3 is confirmed across several frequencies, with differences reaching the significance threshold.



As shown in Figure 3a, word recognition performance improved with amplification. Switching from M0 to M2 significantly increased the WRS by 4.38 percentage points under the 40 dB SPL quiet condition ($p < 0.05$) and by 3.16 percentage points under the 50 dB SPL quiet condition ($p < 0.05$). The improvement was substantially greater in noise, where WRS increased by 11.09 percentage points ($p < 0.001$), indicating a particularly marked enhancement of speech intelligibility under the most challenging listening conditions.

For sentence recognition (Figure 3b), the SRS increased significantly between M0 and M2 under the 40 dB SPL quiet condition, with a gain of 2.73 percentage points ($p < 0.05$). In noise, SRS improved significantly from M0 to M2, with a gain of 16.40 percentage points ($p < 0.001$), and increased further from M2 to M3 by an additional 5.94 percentage points ($p < 0.01$). Overall, this corresponds to a total improvement of 22.34 percentage points between M0 and M3 ($p < 0.001$).

c) Unsupervised Classification (PCA and K-means Clustering):



The principal component analysis showed that the first two principal components accounted for 92.1% of the total variance, with PC1 explaining 83.4% and PC2 explaining 8.7%.

The first principal component (PC1) primarily represented a gradient opposing hearing thresholds (PTA) and speech recognition performance (WRS), with similar standardized loadings for the four variables (absolute values ranging from 0.48 to 0.51). Participants located on the right-hand side of the PCA plot therefore exhibited more severe hearing loss and poorer speech recognition performance.

The second principal component (PC2) explained only a small proportion of the total variance and was primarily driven by the PTA of the worse ear (standardized loading = 0.794) and WRS in quiet (standardized loading = 0.568), thus making a more modest contribution to participant separation.

The standardized coefficients (z-score loadings) associated with each principal component are presented below:

$$PC_1 = 0.514 \times z(PTA_{best}) + 0.478 \times z(PTA_{worst}) - 0.498 \times z(WRS_{quiet,RAU}) - 0.510 \times z(WRS_{noise,RAU})$$

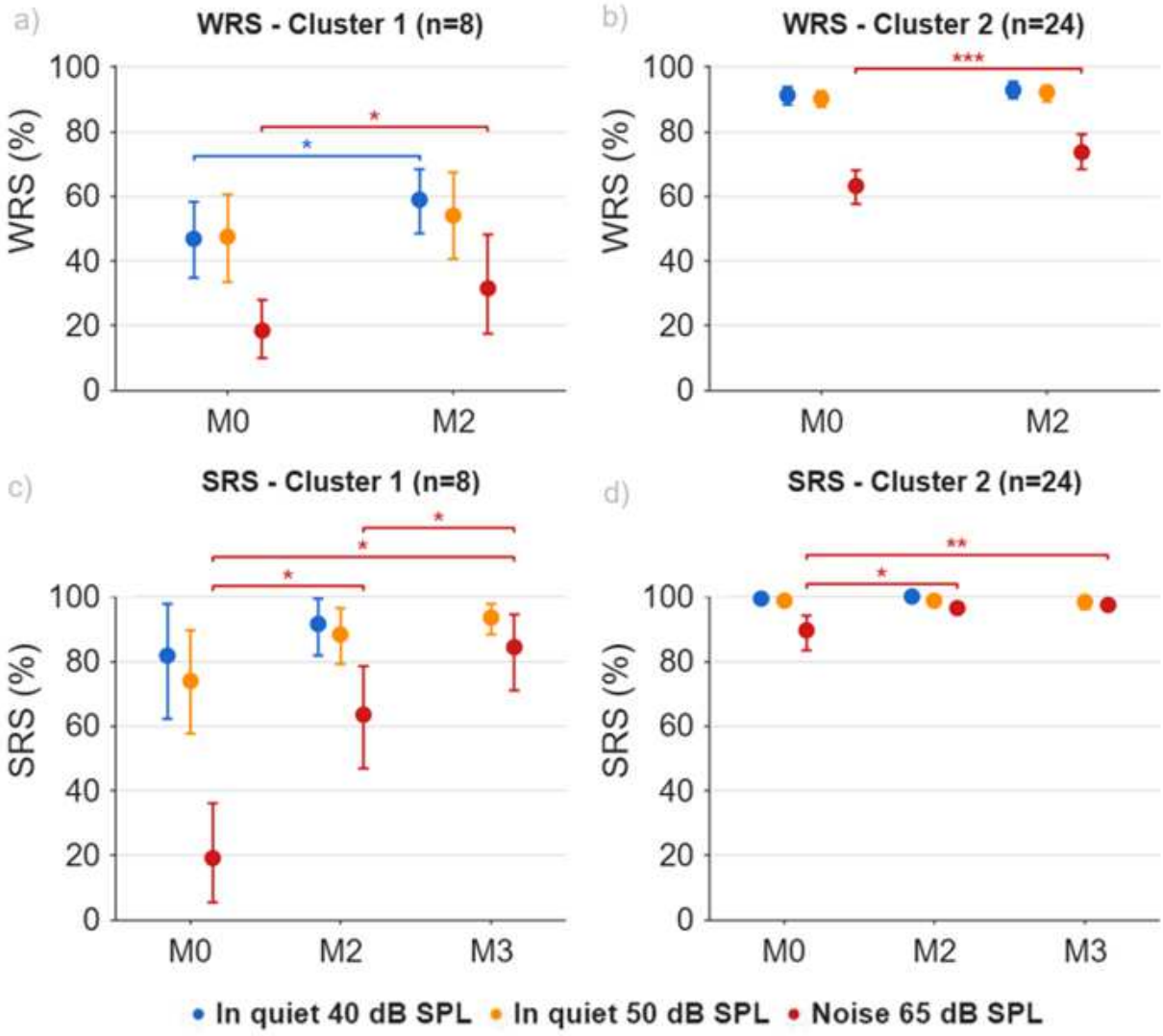
$$PC_2 = 0.026 \times z(PTA_{best}) + 0.794 \times z(PTA_{worst}) + 0.568 \times z(WRS_{quiet,RAU}) + 0.215 \times z(WRS_{noise,RAU})$$

The distribution of participants in the PCA factor space revealed two distinct groups identified by K-means clustering based on the four standardized variables. The selection of a two-cluster solution (k = 2) was supported by both the elbow method and the silhouette coefficient.

Cluster 1 (*n* = 8, 25.0%) comprised participants with more severe hearing loss, with a mean PTA of 43.4 ± 7.5 dB HL for the better ear and 50.6 ± 8.9 dB HL for the worse ear. These participants also exhibited poorer speech recognition performance, with mean WRS values of 47.5 ± 20.3% in quiet and 18.5 ± 13.5% in noise.

In contrast, Cluster 2 (*n* = 24, 75.0%) included participants with better hearing sensitivity (PTAbest: 20.6 ± 8.4 dB HL; PTAworst: 27.5 ± 12.7 dB HL) and substantially better speech recognition performance (WRS in quiet, M0: 90.2 ± 6.4%; WRS in noise, M0: 63.1 ± 13.3%).

d) Effect of Amplification According to Cluster:



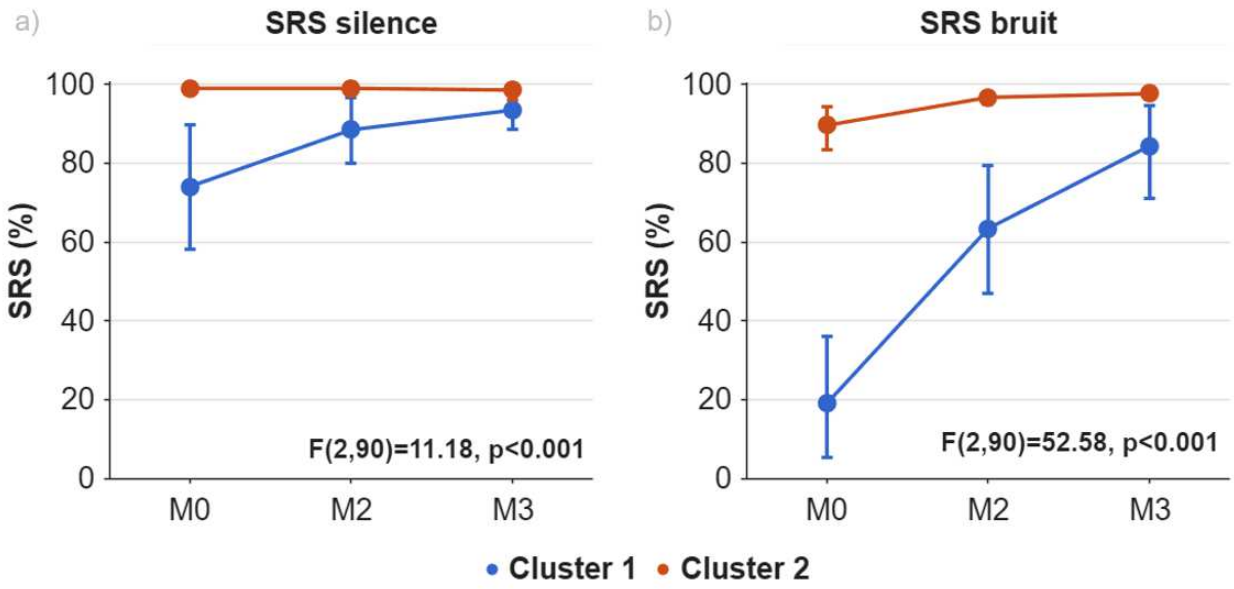
Analysis of WRS and SRS according to cluster revealed distinct response patterns between the two auditory profiles.

Participants in Cluster 1 (n = 8; Figures 5a and 5c), characterized by more severe hearing loss, showed a significant improvement in WRS between M0 and M2 under the 40 dB SPL quiet condition, with a gain of 12 percentage points ($p < 0.05$), as well as in noise, with a gain of 13.5 percentage points ($p < 0.05$).

Sentence recognition also improved significantly, although only under the noise condition. SRS increased by 44.34 percentage points between M0 and M2 ($p < 0.05$) and by 65.34 percentage points between M0 and M3 ($p < 0.05$). Furthermore, the additional improvement of 21 percentage points observed between M2 and M3 demonstrated a significant advantage of the M3 amplification mode.

Participants in Cluster 2 (*n* = 24; Figures 5b and 5d), who exhibited relatively high baseline performance, showed a significant improvement only for WRS in the noise condition. Compared with M0, the M2 mode increased WRS by 10.41 percentage points ($p < 0.001$).

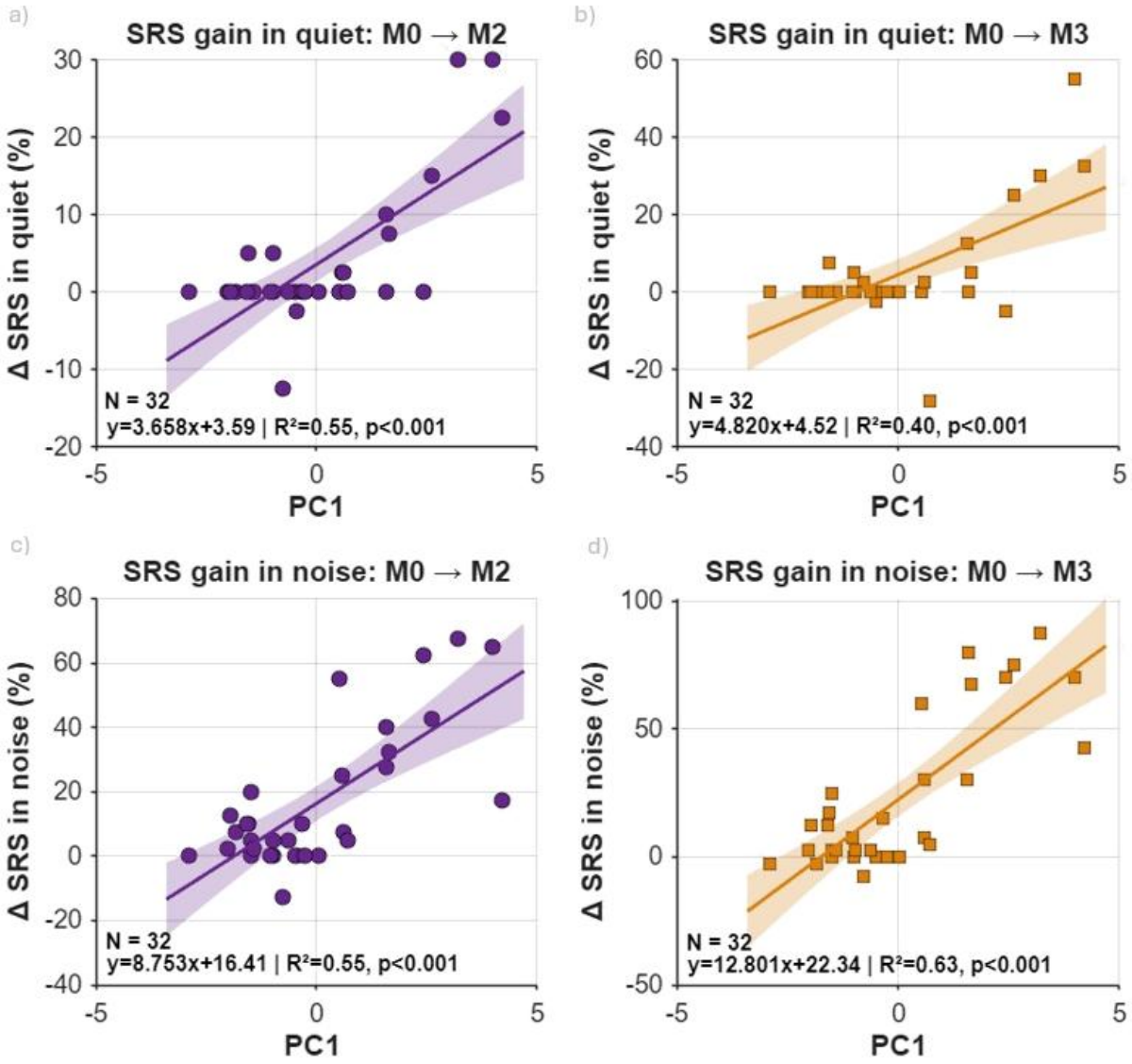
Similarly, significant improvements in sentence recognition were observed only in noise. SRS increased by 7.07 percentage points between M0 and M2 ($p < 0.05$) and by 8.00 percentage points between M0 and M3 ($p < 0.01$), although overall performance remained high across all listening conditions.



To determine whether the benefit provided by amplification differed according to cluster, a linear mixed model including the Mode × Cluster interaction was fitted, with a random subject effect to account for repeated measures (figure 6). This approach allows for a direct test of whether the progression of the score according to amplification mode follows a different trajectory depending on cluster membership.

The Mode × Cluster interaction proved to be highly significant, both in quiet ($F(2,90) = 11.18$, $p < 0.001$, figure 6a) and in noise ($F(2,90) = 52.58$, $p < 0.001$, figure 6b), indicating that the SRS trajectory according to amplification mode differs statistically between the two clusters. This effect is particularly pronounced in noise, where Cluster 1 progresses from 19% to 84.3% between M0 and M3, while Cluster 2 remains nearly stable at around 90-98%.

e) Predictive Model of Individual Amplification Benefit:



Across all four listening conditions, PC1 was a significant predictor of the improvement in SRS provided by amplification ($p < 0.001$ for all models). The explanatory power of the regression models was greater under noisy than quiet listening conditions ($R^2 = 0.55$ for M2 and $R^2 = 0.63$ for M3 in noise, compared with $R^2 = 0.55$ for M2 and $R^2 = 0.40$ for M3 in quiet). In addition, the predictive performance was higher for the M3 mode than for M2 in the noise condition. All four regression models exhibited a positive slope, indicating that participants with greater overall hearing impairment (higher PC1 values) derived larger improvements in sentence recognition from the amplification provided by the soundbar.

f) Questionnaire Results:

The comparisons are summarized in Table 1.

When considering the entire study cohort, the M3 mode received significantly higher ratings than the M0 mode for all six questionnaire items. Likewise, the M2 mode was rated significantly higher than M0 for Items 1, 2, 3, and 6.

When analysed by cluster, participants in Cluster 1 assigned significantly higher scores to M3 than to M0 for Items 1, 2, and 4. For Item 4, M2 also received significantly higher ratings than M0. No statistically significant differences between listening modes were observed for the remaining questionnaire items.

Among participants in Cluster 2, the M3 mode received significantly higher ratings than M0 for all six questionnaire items. In addition, the M2 mode was rated significantly higher than M0 for Items 1, 2, and 3.

IV) Discussion and Limitations:

The overall results of the speech recognition tests show that the benefit provided by the different amplification modes is greater when listening conditions are more difficult (in line with existing literature [18]). The gains observed are highest and most significant, for both WRS and SRS, in the presence of noise at 65 dB SPL. In contrast, improvements are more modest in silence. However, it should be kept in mind that performance in M0 in silence is already very high. Despite the RAU transformation of scores, a ceiling effect probably limited the magnitude of gains observable in this condition: this transformation corrects the variance distortion related to a score's position on the scale, but it cannot compensate for the lack of room for improvement. A participant whose scores remain identical or very close from one mode to another (a subject at 100% in M0, M2, and M3) will always show a gain of zero, regardless of the scale used. This ceiling effect, specific to the already high performance levels in silence, should be kept in mind when interpreting several of the results that follow.

Regarding SRS in silence at 50 dB SPL, the gains between M0 and M2 as well as between M0 and M3 do not reach the threshold of statistical significance, whereas this threshold is reached between M0 and M2 in silence at 40 dB SPL. This difference is explained by the application of a Bonferroni correction: in the 50 dB SPL condition, three modes (M0, M2, M3) are compared with each other, versus only two (M0, M2) in the 40 dB SPL condition. This correction makes the significance threshold stricter as the number of comparisons increases, thereby reducing the likelihood of detecting a statistically significant difference. This same mechanism will come up again later to explain a comparable result regarding the subjective questionnaire.

We also compared our measured audiometric gains with those announced by the manufacturer and observed notable discrepancies. To illustrate, the manufacturer promises approximately 15 dB of gain with the M3 mode at 2000 Hz (Supplementary Fig. S1). However, in our results, for this frequency and this mode, we obtained an average of 8.4 dB (figure 2b). We then noticed that this gain depended on the degree of hearing loss: the more pronounced the hearing loss, the greater the observed gain. We first considered that this might stem from the compression built into the soundbar. However, its activation threshold was never reached, and its activation would have produced the opposite effect (a greater reduction in gain among the most severely

affected participants). After investigation, it turned out that the 40 dB SPL background noise in the booth was actually responsible for reducing our gains. This level is admittedly considered to be silence, but some participants with good hearing are able to detect sounds below 40 dB SPL. The residual noise therefore truncated their results and artificially lowered the device's average gains. For the warble pure tones presented at very low levels to these participants, it was necessary to increase the stimulation level not because of a hearing deficit, but because these stimuli were masked by the residual noise in the booth. Assuming an ideal booth at 0 dB SPL, we estimate that the average gain at 2000 Hz would be closer to the 15 dB announced by the manufacturer (as was the case for the hearing-impaired participants). We conducted a supplementary test with normal-hearing participants, varying the distance to the soundbar. We observed a variation in gain depending on distance, which supports this explanation.

The PCA reveals that a single axis (PC1) is enough to summarize 83.4% of the total variance, opposing hearing thresholds against speech recognition performance. The fact that the four variables contribute almost equally to it confirms that this axis captures a single, coherent dimension ("overall auditory severity"). This result confirms the presence of distinct auditory profiles within the sample, which partly justifies our division into two clusters for subgroup analysis. The second component (PC2) is mainly based on the PTA of the worse ear (coefficient = 0.794, accounting for around 63% of the variance of this axis), with a non-negligible secondary contribution from WRS in silence (coefficient = 0.568, around 32%). The fact that these two coefficients point in the same direction suggests that this axis distinguishes profiles marked by severe impairment of the weaker ear despite good preservation of recognition in silence, rather than reflecting simple overall auditory asymmetry. Nevertheless, it is important to remember that PC2 represents a much more marginal share of the overall data structure (8.7% of variance explained).

The cluster analysis confirms and refines the overall results: the benefit of M2/M3 remains particularly marked under degraded listening conditions, and the observed raw gains vary considerably in magnitude depending on auditory profile. Cluster 1 shows very large gains in SRS in noise (up to +65.34 points between M0 and M3), while Cluster 2 (whose initial performance was already high) shows more modest, though significant, gains (up to +8 points). These raw differences suggest that the benefit of amplification increases with the severity of hearing loss. However, the mere observation that both clusters progress significantly (each considered on its own), with a different magnitude (Figure 5), is not sufficient to conclude that there is a genuine difference in benefit between the two groups [19]. Two complementary approaches were therefore used to settle this question.

The protocol includes three repeated measures per participant (M0, M2, and M3). A linear mixed model including the Mode × Cluster interaction was therefore favored to assess these differences in evolution between clusters. This approach makes use of the full set of repeated SRS measurements rather than a single gain score per subject, which gives it greater statistical power and greater flexibility for longitudinal data analysis, notably through its ability to model within-subject correlation via a random effect [20].

The selected model, including the Mode × Cluster interaction, reveals a highly significant interaction, both in silence ($F(2,90) = 11.18$, $p < 0.001$) and in noise ($F(2,90) = 52.58$, $p < 0.001$). The higher F statistic in noise suggests that the effect of mode depends more strongly on cluster in this condition. This difference should, however, be interpreted with caution, as Cluster 2's scores at M0 were already close to the ceiling of the scale

(see the ceiling effect described above); their margin for improvement was therefore structurally limited. This may artificially inflate the magnitude of the observed interaction.

To explore this relationship from a complementary angle, a regression analysis (figure 7) modeled the association between auditory profile (PC1, used as a continuous index) and SRS gain. PC1 emerged as a significant predictor in all four conditions tested ($p < 0.001$), with a positive linear relationship between the severity of the auditory profile and the benefit of amplification, although the explanatory power of the model remained moderate (R^2 between 0.40 and 0.63). Auditory profile therefore contributes significantly to the observed gain without being its sole determinant. This approach suggests that the benefit of amplification increases progressively with the severity of the auditory profile. A continuous relationship appears more appropriate than a step effect between clusters, while still being partly explained by other individual factors. Finally, it should be noted that, whichever interpretation is adopted, this effect is likely underestimated at the level of the overall sample, since the most severely affected participants represent a minority (25%).

Regarding the questionnaire results, mode M3 does not appear to be significantly preferred over M0 for item 6 ("I prefer this listening mode") among Cluster 1 participants, even though the average scores given to M3 remain higher than those for the other two modes. This result is worth highlighting, since these are precisely the participants who benefit the most from amplification on an objective level. As with SRS in silence at 50 dB SPL discussed above, this lack of significance is likely explained by the Bonferroni correction, further compounded here by the small size of Cluster 1 ($n = 8$), which limits the ability to detect differences that are nonetheless clinically relevant. Conversely, Cluster 2 participants express a significant preference for the amplified modes, even though their (near-normal) auditory profile makes them less reliant on amplification benefit compared to Cluster 1. This favorable subjective acceptance could be explained by how the soundbar distributes gain across frequencies, mainly amplifying high frequencies while limiting gain in low frequencies. The additional amplification therefore does not create discomfort, as the low frequencies responsible for the impression of sound power remain only slightly amplified.

The main limitation of this study concerns the number of subjects included. A large number of participants registered on the lists to take part in studies have normal hearing. It is impossible to know their audiometric profile in advance of recruitment, which made recruiting a varied range of auditory profiles particularly time-consuming. This reduced sample size limits the statistical power of the study and does not allow certain differences to be demonstrated, despite trends suggesting such differences, both for intelligibility performance across listening modes and for questionnaire scores. It would therefore be worthwhile to continue this work by including a larger number of participants.

V) Conclusion:

This study aimed to evaluate the effectiveness of a soundbar incorporating a hearing-aid algorithm based on the NAL-NL2 prescription, designed to improve speech intelligibility in older hearing-impaired adults. The results confirm this objective: modes M2 and M3 significantly improve speech recognition compared with the standard mode, with a particularly marked benefit in degraded listening conditions. The analyses conducted also suggest that this benefit tends to be greater among participants with the most severe auditory profiles.

The regression carried out on the composite index PC1 is consistent with this observation, showing a progressive increase in benefit according to the severity of the auditory profile, rather than a simple threshold effect between clusters.

Beyond objective performance, the questionnaire results also indicate an overall favorable appreciation of the amplified modes, with a significant preference for M3 over M0 at the level of the whole group. Taken together, these results highlight the potential of this technology as a promising solution for improving the television listening experience of older hearing-impaired adults.

Nevertheless, the limited sample size, particularly for the subgroup of the most severely affected participants (n = 8), and the conservative nature of the statistical corrections applied call for some caution in interpreting certain results. Further work involving a larger number of participants will be needed to confirm these observations and to better characterize the user profiles most likely to benefit from these listening modes.

Finally, this study relied mainly on speech intelligibility measures and subjective assessments collected from participants. A particularly interesting direction would be to explore the impact of the different amplification modes on listening effort using physiological measures such as pupillometry. Indeed, if modes M2 and M3 significantly improve speech intelligibility, they could also reduce the cognitive load required to process auditory information. Analysis of pupil diameter variations could complement our results by helping to better understand the link between improved intelligibility and reduced listening effort.

Data availability

The de-identified data supporting the findings of this study are available in Zenodo at https://doi.org/10.5281/zenodo.23081053.

References

1. Gates GA, Mills JH. Presbycusis. Lancet. 2005;366(9491):1111-20. doi:10.1016/S0140-6736(05)67423-5.

2. Statistics Korea (KOSTAT). Population by age group [Internet]. Daejeon: Korean Statistical Information Service (KOSIS); 2025 [cited 9 Sep 2026]. Available from: https://kosis.kr/eng/

3. Cavanaugh JC. Comprehension and retention of television programs by 20- and 60-year olds. J Gerontol. 1983;38(2):190-196.

4. Meyer C, Hickson L, Khan A, Walker D, Dillon H, Hartley D. The benefits of hearing aids and closed captioning for television viewing by older adults with hearing loss. Ear Hear. 2010;31(2):234-241.

5. Anovum GmbH. APAC Trak Korea 2024 [Internet]. Zurich: Anovum GmbH; 2024 [cited 9 Sep 2026]. Available from: https://www.ehima.com/wp-content/uploads/2025/01/KoreaTrak-2024.pdf

6. McClannahan K, Backer KC, Tremblay KL. Auditory Evoked Responses in Older Adults with Normal Hearing, Untreated, and Treated Age-Related Hearing Loss. Ear Hear. 2019;40(5):1106-1116.

7. Keidser G, Dillon H, Flax M, Ching T, Brewer S. The NAL-NL2 Prescription Procedure. Audiol Res. 2011;1(1):e24. doi:10.4081/audiores.2011.e24.

8. Keidser G, Dillon H, Carter L, O'Brien A. NAL-NL2 Empirical Adjustments. Trends Amplif. 2012;16(4):211-223. doi:10.1177/1084713812468511.

9. Kim JS, Lim DH, Hong HN, Shin HW, Lee KD, Hong BN, et al. Development of Korean standard monosyllabic word lists for adults. Korean J Audiol. 2008;12(1):22-28.

10. Lee HJ, Kong TH, Park KH. A Cross-Sectional Study of Word Recognition and Cognitive Function in Older Adults with Hearing Loss Using a Standardized Neuropsychological Battery. J Clin Med. 2024;14(22):7897.

11. Jang HS, Lee JH, Lim DH, Lee KW, Jeon AR, Jung EJ. Development of Korean Standard Sentence Lists for Sentence Recognition Tests. Korean Journal of Audiology. 2008;12(2):178-187.

12. Cox RM, Alexander GC. The International Outcome Inventory for Hearing Aids (IOI-HA): psychometric properties of the English version. Int J Audiol. 2002;41(1):30-35.

13. Cox RM, Alexander GC. The Abbreviated Profile of Hearing Aid Benefit. Ear Hear. 1995;16(2):176-186.

14. World Health Organization. World report on hearing [Internet]. Geneva: World Health Organization; 2021 [cited 9 Sep 2026]. Available from: https://www.who.int/publications/i/item/9789240020481

15. Studebaker DH. A "rationalized" arcsine transform. J Speech Hear Res. 1985;28(3):455-462. doi:10.1044/jshr.2803.455.

16. Pichora-Fuller MK, Schneider BA, Daneman M. How young and old adults listen to and remember speech in noise. J Acoust Soc Am. 1995;97(1):593-608.

17. Olsen WO, Van Tasell DJ, Speaks CE. Phoneme and word recognition for words in isolation and in sentences. Ear Hear. 1997;18(3):175-188. doi:10.1097/00003446-199706000-00001.

18. Horwitz AR, Ahlstrom JB, Dubno JR. Factors affecting the benefits of high-frequency amplification. J Speech Lang Hear Res. 2008;51(3):798-813. doi:10.1044/1092-4388(2008/057).

19. Gelman A, Stern H. The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant. Am Stat. 2006;60(4):328-331. doi:10.1198/000313006X152649.

20. Gueorguieva R, Krystal JH. Move Over ANOVA: Progress in Analyzing Repeated-Measures Data and Its Reflection in Papers Published in the Archives of General Psychiatry. Arch Gen Psychiatry. 2004;61(3):310-317. doi:10.1001/archpsyc.61.3.310.

Acknowledgements: We thank Sin Lyul Lee, Chan Ill Park, Young Ho Je, and Jong Sub Han of JD Solution for providing the soundbar device used in this study and for financial support through a research grant.

Author contributions: M.K.P. conceived and supervised the study. C.Y. and G.D.S. collected and analysed the data. G.D.S. drafted the manuscript, and M.K.P. critically revised it. All authors reviewed and approved the final version of the manuscript.

Additional Information

Competing interests: This work was supported by a research grant from JD Solution, the manufacturer of the soundbar evaluated in this study, which also provided the device on loan. JD Solution had no role in the study design, data collection, data analysis, interpretation of the data, or preparation of the manuscript. The authors declare no other financial or non-financial competing interests.

Figure legends

Figure 1. Baseline audiological characteristics of the 32 participants. (a) Mean pure-tone audiogram (dB HL) as a function of frequency (Hz), with the mean Pure-Tone Average (PTA) shown in the inset. (b) Speech Recognition Threshold (SRT; 66% criterion). (c) Word Recognition Score at the Most Comfortable Level (WRS-MCL). Error bars represent 95% confidence intervals. Comparisons between the better and worse ears were performed using the Wilcoxon signed-rank test *** : $p < 0.001$; ns : not significant).

Figure 2. Free-field pure-tone thresholds and hearing threshold improvements according to amplification mode in the 32 participants. (a) Mean free-field hearing thresholds (dB HL) obtained using the M0, M2, and M3 amplification modes as a function of frequency (Hz). (b) Mean threshold improvements (dB) relative to the M0 mode for M2 and M3 across frequencies. Error bars represent 95% confidence intervals. Comparisons were performed using the Wilcoxon signed-rank test with Bonferroni correction (* : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$; ns : non-significant). The row labelled "M2 vs. M3" indicates frequencies at which significant differences were observed between the two amplification modes.

Figure 3. Speech recognition performance as a function of listening condition (N = 32). (a) Word Recognition Score (WRS, %) measured in quiet (40 and 50 dB SPL) and in noise (65 dB SPL) using the M0 and M2 amplification modes. (b) Sentence Recognition Score (SRS, %) measured under the same listening conditions using the M0, M2, and M3 modes. Data are presented as mean ± 95% confidence interval (percentile bootstrap). Statistical comparisons between amplification modes were performed using the Wilcoxon signed-rank test on RAU-transformed scores (* : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$). A Bonferroni correction was applied to comparisons involving the three amplification modes.

Figure 4. Distribution of participants in the principal component analysis (PCA) factor space based on four standardized variables: Pure-Tone Average (PTA) of the better ear (PTAbest), PTA of the worse ear (PTAworst), Word Recognition Score (WRS) in quiet using the M0 mode, and WRS in noise using the M0 mode. WRS values were transformed into Rationalized Arcsine Units (RAU) prior to analysis. Clusters were identified using the K-means algorithm applied to the PCA scores. The optimal number of clusters (k = 2) was determined using the elbow method (inset), based on the within-cluster sum of squares (WCSS). Arrows indicate the loadings of the original variables on the first two principal components.

Figure 5. Speech recognition performance according to amplification mode by cluster (N = 32). (a) Word Recognition Score (WRS, %) for Cluster 1 (n = 8) in quiet (40 and 50 dB SPL) and in noise (65 dB SPL) using the M0 and M2 modes. (b) WRS for Cluster 2 (n = 24) under the same listening conditions. (c) Sentence Recognition Score (SRS, %) for Cluster 1 using the M0, M2, and M3 modes. (d) SRS for Cluster 2 under the same listening conditions. Data are presented as mean ± 95% confidence interval (percentile bootstrap). Statistical comparisons between amplification modes were performed using the Wilcoxon signed-rank test on RAU-transformed scores (* : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$). A Bonferroni correction was applied for comparisons involving the three amplification modes.

Figure 6. Sentence Recognition Score (SRS, %) according to amplification mode (M0, M2, M3) and cluster membership, in quiet and in noise. (a) SRS in quiet (50 dB SPL). (b) SRS in noise (65 dB SPL). Data are expressed as mean ± 95% confidence interval (percentile bootstrap). Cluster 1 (blue, n = 8) comprises participants with more severe hearing loss, Cluster 2 (red, n = 24) comprises participants with better hearing. The Mode × Cluster interaction was tested using a linear mixed model (Score ~ Mode × Cluster + (1|Subject)), allowing the within-subject repeated-measures structure to be accounted for. SRS scores were transformed into RAU units prior to analysis.

Figure 7. Linear regression analysis between the composite hearing severity index (PC1), derived from the principal component analysis, and the improvement in Sentence Recognition Score (ΔSRS, %) provided by the M2 and M3 amplification modes relative to M0 (N = 32). (a) SRS improvement in quiet for M2 relative to M0. (b) SRS improvement in quiet for M3 relative to M0. (c) SRS improvement in noise for M2 relative to M0. (d) SRS improvement in noise for M3 relative to M0. Each point represents one participant. The solid line represents the fitted linear regression, and the shaded area indicates the 95% confidence interval of the predicted values.

Tables

Table 1. Results of the six-item questionnaire evaluating the three listening modes (scores ranging from 0 to 10). Statistical comparisons were performed using the Wilcoxon signed-rank test with Bonferroni correction. Inequalities indicate that one listening mode received a significantly higher or lower score than another ($p < 0.05$). n.s.: not significant.

Group	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Global	M0<M2<M3	M0<M2, M0<M3	M0<M2, M0<M3	M0<M3	M0<M3	M0<M2, M0<M3
Cluster 1	M0<M3	M0<M3	n.s	M0<M2, M0<M3	n.s	n.s
Cluster 2	M0<M2, M0<M3	M0<M2, M0<M3	M0<M2, M0<M3	M0<M3	M0<M3	M0<M3

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SoundbarSupplementaryInformation.pdf