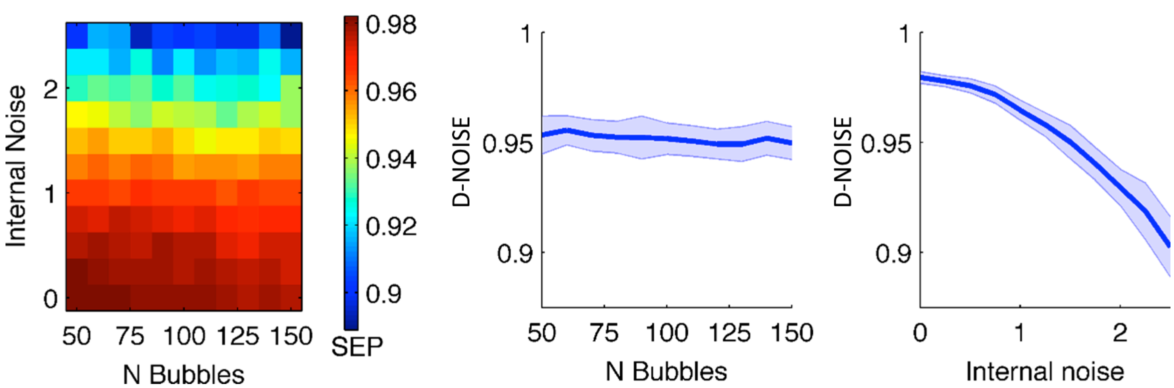
Supplemental Observer Simulation

1. Ideal Observer Simulation

We performed a simulation in which a simulated listener performed the same experimental procedure as the real listeners at each of several values for average number of bubbles at threshold (50–150 in steps of 10, the range observed for real older hearing-impaired [OHI] listeners). The simulated listener was assumed to generate a distribution of responses (keywords correct) that matched the average distribution of responses for the real OHI listeners. Responses for each trial were generated by observing a decision variable that reflected the degree of overlap between the bubbles masker and an idealized CImg (average of young normal-hearing [YNH] listeners), and sorting this variable across trials such that the largest values were assigned a response of 5, the next-largest values were assigned a response of 4, and so on, until the smallest values were assigned a response of 0, always respecting the assumed distribution of responses. Classification image analysis was then performed using the weighted sum technique of Venezia, Hickok, and Richards (2016) to minimize computational load, and a value of D-NOISE was obtained for each simulation run. In addition to varying the average number of bubbles, the degree of internal noise was orthogonally varied by adding different amounts of Gaussian noise to the decision variable at each value for the average number of bubbles. The experiment was simulated 100 times for each possible combination of average number of bubbles and internal noise. We then compared the effect of increasing the number of bubbles to the effect of increasing the internal noise. Supplemental Figure 1, below, clearly shows that increasing the number of bubbles did not produce a change in D-NOISE, while increasing the level of internal noise systematically caused D-NOISE to decrease. The range of simulated D-NOISE values across different levels of internal noise was also similar to the range of true D-NOISE values observed for OHI listeners (Figure 7 of main article). There was no apparent interaction between number of bubbles and internal noise.



Supplemental Figure 1. Ideal-observer simulation comparing effects of number of bubbles and internal noise on D-NOISE. The simulated listener performed 100 replications of the experiment at fully crossed combinations of number of bubbles (N bubbles) and Internal Noise. Responses for each trial were generated by observing a decision variable (DV) that reflected the degree of overlap between a randomly-generated bubbles masker and an idealized classification image (z-scored average of YNH listeners). The overlap calculation was restricted to regions of the idealized classification image for which z > 2. A null distribution on the DV was generated by repeatedly observing the DV using different randomly generated maskers with the assumed number of bubbles, and a criterion value was then set as the median of the null distribution. In a single simulation run, the simulated listener performed the experiment following same the up-down adaptive rule imposed on real listeners. The criterion value for the DV was used to establish correct vs. incorrect for the adaptive rule. Gaussian noise was added to the DV with zero mean *and standard deviation equal to a scaling factor (Internal Noise) multiplied to the standard deviation of the null distribution of the DV. A CImg was generated after each simulation run and D-NOISE was calculated. Left: mean (across 100 simulation runs) value of D-NOISE across the range of N Bubbles (abscissa) and Internal Noise (ordinate). Middle: marginal distribution of D-NOISE (ordinate) across values of N Bubbles (abscissa). The mean is given by the bold line and the standard deviation is given by the shaded region around the line. Right: marginal distribution of D-NOISE (ordinate) across values of Internal Noise (abscissa). The mean is given by the bold line and the standard deviation is given by the shaded region around the line.*

2. Coherence Speech Intelligibility Index (CSII)

Normally, the speech intelligibility index (SII; American National Standards Institute [ANSI] S3.5-1997) is calculated by separately estimating the equivalent spectrum levels of a target speech signal and an additive noise signal. In the present study, we have speech signals that have been distorted such that we cannot disentangle “noise” from “speech plus noise” in the distorted signal. Kates and Arehart (2005) developed a method to calculate the SII in the presence of distortion using the magnitude squared coherence (MSC), which is the squared magnitude of the normalized cross-spectral density between two signals. Their method, known as the coherence speech intelligibility index (CSII), assumes that, given a clean input speech signal, $x(n)$ and a distorted output signal, $y(n)$, the MSC represents the fraction of the signal power in $y(n)$ that is linearly related to $x(n)$ at each analysis frequency. The complementary value, 1-MSC, gives the fraction of the output power that is not linearly related to the input and thus represents noise and nonlinear distortion. Therefore, the power in the speech signal can be calculated by weighting the power spectral density of $y(n)$ by the MSC and summing across analysis frequencies. The noise power can be calculated in the same way but weighting by 1-MSC. To calculate the equivalent speech and noise spectra, speech and noise power are calculated separately across critical bands with center frequencies and bandwidths as specified in ANSI S3.5-1997. The output power of each band is then converted to a spectrum level based on the specified bandwidth. We calculated the equivalent speech and noise spectra using this technique and then submitted these values to the critical band procedure for calculating the SII (ANSI S3.5-1997). Participants’ audiograms were interpolated to critical band center frequencies using linear interpolation on the log scale and converted to equivalent internal noise such that effects of audibility were included in the calculation.

The resulting metric is the CSII, which ranges from 0–1 and represents the proportion of speech information transmitted to the listener. The CSII therefore jointly captures the effects of audibility and nonlinear distortion. Following Kates and Arehart (2005), we calculated the CSII separately for low, medium, and high amplitude portions of the speech envelope. This was performed by windowing the clean speech signal (hamming window, 16 ms block size, 50% overlap), calculating the magnitude of the signal in each segment, and comparing the segment magnitude to the overall root-mean-squared (RMS) level of the segments. High-level segments were those at or above the RMS level, mid-level segments were 0-10 dB below the RMS level, and low-level segments were 10–30 dB below the RMS level. The samples belonging to each segment were concatenated separately for high-, mid-, and low-level segments, and these concatenated signals were submitted separately for calculation of the CSII. The high-, mid-, and low-amplitude estimates of the CSII were averaged to create a single metric that was used to predict across-trial threshold performance.

The across-trial average CSII was significantly different across listener groups (Welch’s F test, $p < .001$). Specifically, there were significant differences between the YNH group and the older groups (vs. ONH: $t(26.0) = 6.7$, $p < .001$; vs. OHI: $t(22.5) = 3.8$, $p < .01$) but not between the OHI and ONH groups ($t(29.6) = 0.21$, $p > .9$). The YNH listeners had a lower CSII than the older normal-hearing (ONH)/OHI groups, indicating better performance (more stimulus distortion tolerated = lower CSII = better performance). Therefore, when audibility is accounted for, there is no difference in threshold performance between the ONH and OHI groups (i.e., CSII is not significantly different between these groups), but a significant difference between the YNH and ONH/OHI groups remains. This suggests that audibility alone does not account for the performance difference between younger and older listeners.

References

American National Standards Institute. (1997). *Methods for the calculation of the speech intelligibility index (S3.5-1997)*. New York, NY: Author.

Kates, J. M., & Arehart, K. H. (2005). Coherence and the speech intelligibility index. *The Journal of the Acoustical Society of America, 117,* 2224–2237.

Venezia, J. H., Hickok, G., & Richards, V. M. (2016). Auditory “bubbles”: Efficient classification of the spectrotemporal modulations essential for speech intelligibility. *The Journal of the Acoustical Society of America, 140,* 1072–1088.