

Research Article

Identification of the Spectrotemporal Modulations That Support Speech Intelligibility in Hearing-Impaired and Normal-Hearing Listeners

Jonathan H. Venezia,^{a,b} Allison-Graham Martin,^c Gregory Hickok,^c and Virginia M. Richards^c

Purpose: Age-related sensorineural hearing loss can dramatically affect speech recognition performance due to reduced audibility and suprathreshold distortion of spectrotemporal information. Normal aging produces changes within the central auditory system that impose further distortions. The goal of this study was to characterize the effects of aging and hearing loss on perceptual representations of speech.

Method: We asked whether speech intelligibility is supported by different patterns of spectrotemporal modulations (STMs) in older listeners compared to young normal-hearing listeners. We recruited 3 groups of participants: 20 older hearing-impaired (OHI) listeners, 19 age-matched normal-hearing listeners, and 10 young normal-hearing (YNH) listeners. Listeners performed a speech recognition task in which randomly selected regions of the speech STM spectrum were revealed from trial to trial. The overall amount of STM information was varied using an up-down staircase

to hold performance at 50% correct. Ordinal regression was used to estimate weights showing which regions of the STM spectrum were associated with good performance (a “classification image” or Clmg).

Results: The results indicated that (a) large-scale Clmg patterns did not differ between the 3 groups; (b) weights in a small region of the Clmg decreased systematically as hearing loss increased; (c) Clmgs were also nonsystematically distorted in OHI listeners, and the magnitude of this distortion predicted speech recognition performance even after accounting for audibility; and (d) YNH listeners performed better overall than the older groups.

Conclusion: We conclude that OHI/older normal-hearing listeners rely on the same speech STMs as YNH listeners but encode this information less efficiently.

Supplemental Material: <https://doi.org/10.23641/asha.7859981>

Persons with sensorineural hearing loss often report that talkers seem to be “mumbling” or are otherwise difficult to understand even when the target speech is clearly audible. This difficulty is exacerbated when listening in noise or in the presence of multiple simultaneous talkers (Gatehouse & Noble, 2004). These effects are partially accounted for by loss of sensitivity and the resultant reduction in the effective bandwidth of speech in high-frequency hearing loss. However, hearing loss is also associated with a reduction in suprathreshold auditory function, including

loss of frequency selectivity, poor temporal resolution, and abnormal growth of loudness (cf. Moore, 1996). This two-factor characterization of hearing loss has been useful in explaining the speech perception deficits experienced by hearing-impaired listeners. Indeed, it has long been known that hearing-impaired listeners require a more favorable signal-to-noise ratio to recognize speech in background noise compared to normal-hearing listeners (Cohen & Keith, 1976; Cooper & Cutts, 1971; Kuzniarz, 1973), regardless of whether speech is presented at low or high (suprathreshold) levels (Dirks, Morgan, & Dubno, 1982; Duquesnoy, 1983; Plomp, 1978; Plomp & Duquesnoy, 1982; Plomp & Mimpen, 1979). Plomp (1978, 1986) formalized this concept by showing that the speech reception thresholds of hearing-impaired listeners can be predicted across a range of noise levels using a two-factor model: Parameter A (attenuation), which relates to elevation of pure-tone thresholds and explains the higher speech levels required by hearing-impaired listeners at low noise levels, and Parameter D (distortion), which explains

^aVA Loma Linda Healthcare System, CA

^bDepartment of Otolaryngology, School of Medicine, Loma Linda University, Loma Linda, CA

^cDepartment of Cognitive Sciences, University of California, Irvine

Correspondence to Jonathan H. Venezia: jonathan.venezia@va.gov

Editor-in-Chief: Frederick (Erick) Gallun

Editor: Jennifer Lentz

Received February 6, 2018

Revision received June 5, 2018

Accepted October 16, 2018

https://doi.org/10.1044/2018_JSLHR-H-18-0045

Disclosure: The authors have declared that no competing interests existed at the time of publication.

the higher speech levels required by hearing-impaired listeners at suprathreshold noise levels.

Across listeners, both the A-term and the D-term are related to degree of hearing loss expressed in terms of pure-tone thresholds. However, there is considerable variability in speech reception performance in noise (D) even for listeners who perform similarly in quiet (A+D). That is, distortion is a source of considerable individual variability. Therefore, it is no surprise that audibility accounts for, at maximum, about 40%–50% of the variance in speech recognition performance in hearing-impaired listeners (Humes, 2002). Moreover, restoration of audibility by amplification often does not correct for speech perception difficulties in noisy environments (Kochkin, 2010).

Although suprathreshold distortion clearly affects performance in hearing-impaired listeners, research investigating particular aspects of such distortion—for example, temporal versus spectral resolution—and their relation to speech perception in noise has failed to pin down a particular factor or combination of factors that is consistently able to predict individual differences (cf. Houtgast & Festen, 2008). Further complicating the matter is that speech perception deficits in hearing loss are difficult to separate from age-related speech perception deficits. The majority of persons with hearing loss are older/elderly (Lin, Niparko, & Ferrucci, 2011). Older persons with normal hearing show deficits in temporal processing (Fitzgibbons & Gordon-Salant, 2010; Moore, 2016) and speech perception in noise (Dubno, Dirks, & Morgan, 1984; Füllgrabe, Moore, & Stone, 2015) that resemble the suprathreshold deficits observed in hearing-impaired listeners. However, although deficits in hearing loss can be accounted for in terms of changes in cochlear mechanics caused by outer hair cell loss (Oxenham & Bacon, 2003), age-related deficits likely have a central origin, which may include loss of temporal precision in the auditory brainstem (Anderson, Parbery-Clark, White-Schwoch, & Kraus, 2012), loss of inhibition at various stages of the auditory system (Caspary, Ling, Turner, & Hughes, 2008; Herrmann, Henry, Johnsrude, & Obleser, 2016; Presacco, Simon, & Anderson, 2016; Stothart & Kazanina, 2016), neuronal loss in the auditory cortex (Profant et al., 2014), and changes in auditory cortical connectivity (cf. Cardin, 2016). Moreover, aging affects aspects of cognition that contribute to speech understanding in complex listening environments, including working memory, executive function, and processing speed (Albinet, Boucard, Bouquet, & Audiffren, 2012; Braver & West, 2008; George et al., 2007; Humes & Dubno, 2010; Salthouse, 2000; Schneider, Pichora-Fuller, & Daneman, 2010).

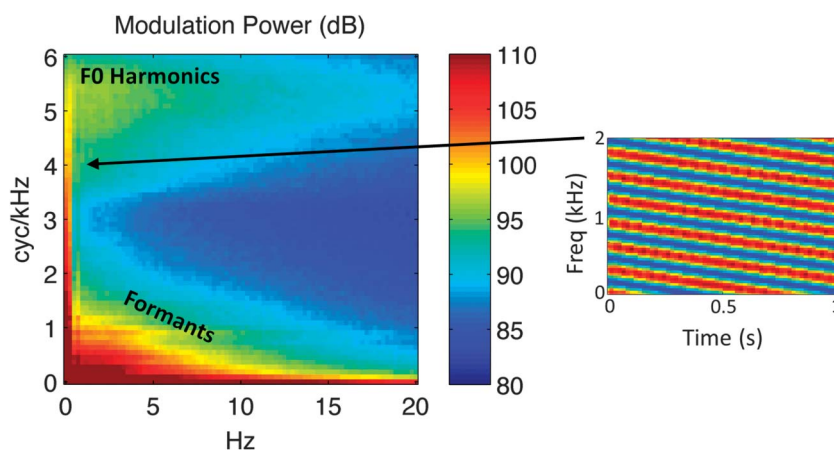
Bernstein (2016) recently described a framework in which spectrotemporal modulation (STM) sensitivity can be used to capture the relation between suprathreshold distortion and speech perception in noise by hearing-impaired listeners. STMs are fluctuations in sound energy across time and frequency. Sinusoidal STMs can be produced by generating rippled noises in which the spectral peak frequencies drift over time. Such noises are characterized by the density of the spectral ripple (spectral modulation rate, in cycles

per octave [cyc/octave] or cycles per kilohertz [cyc/kHz]) and the rate of temporal drift (temporal modulation rate, in hertz [Hz]). Recent work by Bernstein and colleagues demonstrates the following: (a) STM sensitivity (modulation depth at detection threshold) accounts for 40% of the variance in speech-in-noise performance in hearing-impaired listeners, over and above the variance accounted for by audibility (Bernstein et al., 2013); (b) STM sensitivity in hearing-impaired listeners is reduced only for low temporal modulation rates (4–12 Hz) and at low carrier frequencies (< 2 kHz), that is, for spectrotemporal patterns likely to be found in speech (Bernstein et al., 2013; Mehraei, Gallun, Leek, & Bernstein, 2014); and (c) sensitivity to these speech-like STMs accounts for an additional 13% of the variance in speech reception performance in hearing-impaired listeners, over and above the variance accounted for by pure-tone thresholds, even when speech is amplified using an individualized frequency-dependent gain (Bernstein et al., 2016). The authors speculate that STM sensitivity performs so well as a predictor of speech-in-noise performance because it captures important aspects of suprathreshold distortion—namely, temporal fine structure processing and frequency selectivity—as they relate to analysis of complex, speechlike sounds. From a neurobiological perspective, the response of cortical neurons and neuronal ensembles to speech and other complex sounds is well described in terms of STM tuning (Hullett, Hamilton, Mesgarani, Schreiner, & Chang, 2016; Pasley et al., 2012; Shamma, 2001). The auditory cortex has been modeled as an STM filterbank (Chi, Gao, Guyton, Ru, & Shamma, 1999; Chi, Ru, & Shamma, 2005), and the output of such computational models can be used to predict speech intelligibility in the presence of noise or other signal degradations (Elhilali, Chi, & Shamma, 2003). For hearing-impaired listeners, cortical representations of STMs presumably reflect the combined effects of distortion introduced at the auditory periphery and in ascending processing centers of the auditory pathway. As such, STM sensitivity may also be capable of capturing age-related auditory processing deficits.

This study was designed to characterize the perceptual representation of speech STMs in older normal-hearing (ONH) and older hearing-impaired (OHI) listeners, as well as in young normal-hearing (YNH) listeners for comparison. We recently devised a procedure, “auditory bubbles” (Venezia, Hickok, & Richards, 2016), that determines which regions of the two-dimensional (2D) speech modulation spectrum (see Figure 1) contribute most to intelligibility (see also Mandel, Yoho, & Healy, 2016). The speech signal can be described as a weighted sum of STMs with different combinations of spectral and temporal modulation rates (Singh & Theunissen, 2003), and the modulation spectrum—obtained by 2D Fourier transform of the speech spectrogram—shows long-term speech energy across STMs.¹ The auditory

¹STMs with spectral peaks drifting upward versus downward in frequency are typically described separately in the modulation spectrum, but here, we assume symmetry in the distribution of energy across upward- and downward-sweeping STMs.

Figure 1. Average long-term modulation spectrum of 452 sentences spoken by a female talker. The modulation spectrum describes the speech spectrogram as a weighted sum of spectrotemporal ripples containing energy at a unique combination of temporal (abscissa, hertz [Hz]) and spectral (ordinate, cycles per kilohertz [cyc/kHz]) modulation rate. Each pixel represents the speech energy at that particular combination. An example of a downward-sweeping ripple describing a 2-Hz, 4-cyc/kHz STM is shown at the right. In this figure and throughout this article, energy at a given pixel location in the modulation spectrum reflects the average of downward- and upward-sweeping ripples at a given combination of Hz and cyc/kHz. When the modulation spectrum is computed from a speech spectrogram with a linear frequency axis, modulation energy clusters into two discrete regions: a “high spectral modulation rate” region corresponding to finely spaced harmonics of the fundamental (F0 Harmonics) and a “low spectral modulation rate” region corresponding to coarsely spaced resonant frequencies of the vocal tract (Formants). Color scale in decibels (arbitrary reference). Freq = frequency.



bubbles procedure works by randomly removing speech energy in different regions of the modulation spectrum from trial to trial of a sentence recognition task, which allows identification of the regions for which disruption of the signal produces a consistent effect on intelligibility (keywords correctly identified). The outcome is a set of weights—a classification image (CImg) in the 2D modulation spectrum domain—showing which STMs listeners rely on to achieve intelligible perception. Building on the work of others (Drullman, Festen, & Plomp, 1994a, 1994b; Elliott & Theunissen, 2009; Ter Keurs, Festen, & Plomp, 1992, 1993a), we showed that YNH listeners rely primarily on a circumscribed region of the modulation spectrum comprising low spectral (< 2 cyc/kHz) and temporal (< 10 Hz) modulation rates (i.e., STMs within the “formant region” of Figure 1). CImgs had a low-pass character in the spectral dimension and a bandpass character with a peak at ~4 Hz, roughly the syllable rate (Arai & Greenberg, 1997), in the temporal dimension.

Here, we leveraged the ability to measure CImgs to learn about internal distortion of the speech signal at the peripheral or central levels of processing related to aging and/or hearing loss, hypothesizing that such changes would alter/distort the observed CImg. Specifically, we examined two primary CImg measures across YNH, ONH, and OHI listeners: (a) the pattern of CImg weights as described above and (b) the ability of the CImg model to predict individual participant performance from independent data (i.e., data “unseen” by the model). The auditory bubbles method is fundamentally a data-driven technique, so we did not have strong hypotheses with respect to Measure 1 or 2.

Therefore, this should be considered a “hypothesis generating” more so than a “hypothesis testing” study. However, past research provides some direction with respect to potential outcomes. First, regarding Measure 1, it is known that OHI listeners have temporal processing deficits, including difficulty in processing speech information at high temporal rates (Gordon-Salant, 2001). Additionally, hearing-impaired listeners show reduced sensitivity to high-rate spectral modulations (Henry, Turner, & Behrens, 2005). Our initial auditory bubbles study (Venezia et al., 2016) had shown that, although listeners rely primarily on relatively low-rate STMs for intelligibility, they are also capable of using speech information up to about 20 Hz and 4–6 cyc/kHz. The bubbles task forces listeners to rely on a subset of the STM spectrum, which often excludes some or all of the maximally important low-rate STMs. Therefore, if older and/or hearing-impaired listeners are less able to make use of the partially redundant but impoverished high-rate STM cues located toward the edges of peak regions in the CImg, there would be a general shift in the pattern of the CImg toward lower STM rates.

A second potential outcome regarding Measure 1 is based on work suggesting that OHI listeners process speechlike STMs less efficiently than normal-hearing listeners, although not necessarily due to a decrease in spectral or temporal resolution (Bacon & Viemeister, 1985; Grant, Summers, & Leek, 1998; Moore, 2016; Shen, 2014; Ter Keurs, Festen, & Plomp, 1993b). This predicts that ONH and/or OHI listeners may have CImgs that differ within the region of primary importance observed in YNH listeners. That is, changes would not concern which STMs

contribute to intelligibility but rather how well those STMs are encoded during speech processing.

Regarding Measure 2, it is important to note that the pattern of CImg weights (Measure 1) is not necessarily tied to CImg model predictive performance. That is, the CImg model links changes in the speech stimulus (“bubbles” filter patterns) to changes in the response (keywords correctly identified) via a transfer function (CImg weights) that is presumably sensitive to some but not all aspects of aging and hearing loss that could affect performance. The relative (in)ability of the CImg model to predict performance tells us the extent to which such “unexplained” factors contribute to the performance of a given listener. One possibility is that auditory and/or cognitive deficits that do not produce systematic changes in the CImg weights nonetheless produce an increase in the “noise” of the CImg model, detectable as a reduction in model predictive performance.

To assess these potential outcomes, we obtained Measures 1 and 2 for each listener in our three groups. We tested for differences in the listener group means for each measure. In fact, we did not find strong evidence for group differences on Measure 1, although we did find a trend toward an effect of hearing loss, and we found that OHI listeners were significantly different from the YNH/ONH listeners on Measure 2. We therefore devised a number of post hoc analyses to investigate the effect of hearing loss in more detail. We should note that both Measures 1 and 2 are potentially responsive to changes at multiple levels of processing—that is, audibility, peripheral and central suprathreshold distortion, and cognition—all of which are affected in OHI listeners (see above). Therefore, in the post hoc analyses, we were careful to distinguish effects of age from effects of hearing loss per se, and we further decomposed effects of hearing loss in terms of those that could be accounted for by audibility (pure-tone thresholds) versus those that could not.

In addition to measures regarding the CImgs, the results were studied in terms of a measure of overall performance on the auditory bubbles task reflecting the proportion of the original speech modulation spectrum required to achieve 50% correct keyword recognition (threshold). Listener group differences were assessed to determine whether older and/or hearing-impaired listeners would require more STM information (i.e., less stimulus distortion) to perform the task at threshold. This is akin to what Houtgast and Festen (2008) refer to as the “distortion sensitivity approach” to characterizing suprathreshold effects on speech reception. Correlations of threshold performance with age, degree of hearing loss, and CImg summary metrics were examined within the older listener groups.

Method

Participants

This study included 10 YNH (six female, four male) listeners with pure-tone thresholds equal to or less than 20 dB HL at audiometric frequencies from 250 to 8000 Hz ($M_{\text{age}} =$

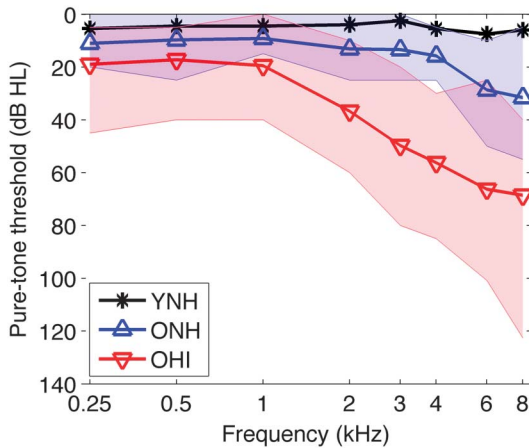
26.9 years, range: 18–34 years), 19 ONH (14 female, five male) listeners with pure-tone thresholds equal to or less than 25 dB HL at audiometric frequencies from 250 to 4000 Hz ($M_{\text{age}} = 66.3$ years, range: 56–78 years), and 20 OHI (eight female, 12 male) listeners with moderate to severe hearing loss who failed to meet the criteria for the ONH group ($M_{\text{age}} = 71.2$ years, range: 51–84 years). An exception to these inclusion criteria was made for one participant in the YNH group who had a pure-tone threshold of 25 dB HL at 6 kHz. The audiometric inclusion criteria were applied to the ear with the lower pure-tone average (PTA) threshold (0.5, 1, and 2 kHz). If the PTA was equal across ears, the criteria were applied to the ear with the lower average threshold across all frequencies tested. The left–right ear difference in the four-frequency PTA (0.5, 1, 2, and 4 kHz) was ≤ 15 dB for all listeners, with the exception of three OHI listeners, two of whom had asymmetries < 25 dB and one of whom had a profound hearing loss in one ear. For all listeners, experimental stimuli were presented monaurally to the better-hearing ear.

Although the ONH and OHI groups were not significantly different in terms of age, $t(37) = 1.81, p = .08$, many of the oldest participants (i.e., > 75 years old) were in the OHI group. Participants were verbally screened by the experimenters for neuropsychological conditions and to determine English fluency. Participants were recruited to participate at one of two testing sites: (a) the Hearing Lab at the University of California, Irvine (17 OHI, 14 ONH, and five YNH; 21 female, 15 male) or (b) the Auditory Research Lab at the VA Loma Linda Healthcare System (three OHI, five ONH, and five YNH; seven female, six male). At Site 1, participants were recruited from the surrounding community via newspaper advertisement; the University of California, Irvine Mind Institute’s C2C registry; or the Social Sciences Human Subjects Lab. At Site 2, participants were recruited from veteran medical record searches, advertisements placed around the VA hospital, or the laboratory’s existing participant database. All participants were compensated \$10/hr for their time, except for five volunteer participants at Site 2. Figure 2 plots the group-averaged audiograms. Thresholds were not obtained at 0.25 kHz for two YNH participants due to experimenter error, and these values were set equal to the 0.5-kHz threshold. Thresholds could not be obtained at 8 kHz for eight OHI participants or at 6 kHz for one OHI participant. These values were interpolated based on fourth-order orthogonal polynomial growth curves fit to the obtained threshold data (Vaden, Matthews, Eckert, & Dubno, 2017).

Stimuli

The stimuli used in this study have been described in detail previously by Venezia et al. (2016). Briefly, the target speech signals were recordings of 452 sentences from the Institute of Electrical and Electronics Engineers sentence corpus (IEEE, 1969) spoken by a single female talker. Each sentence was stored as a separate .wav file digitized at 22,050 Hz with 16-bit quantization. The sound files were

Figure 2. Group-averaged pure-tone air-conduction audiograms. Mean pure-tone thresholds for the young normal-hearing (YNH; black), older normal-hearing (ONH; blue), and older hearing-impaired (OHI; red) listeners are indicated by bold lines. The threshold ranges for the ONH and OHI listeners are indicated by the blue- and red-shaded regions, respectively. The purple-shaded region indicates an overlap in the ONH and OHI ranges. Threshold ranges for the YNH listeners are not shown.



zero-padded to a duration of 3.29 s. To create the experimental stimuli, the sentence audio files were filtered to remove randomly selected regions of energy in the STM domain. For each sentence, a log power (dB) spectrogram was obtained using Gaussian windows with a 4.75 ms–33.5 Hz time–frequency scale. The 2D modulation spectrum was then obtained as the modulus of the 2D Fourier transform of the spectrogram. A 2D filter of the same dimensions as the modulation spectrum was created by generating an identically sized image with a predetermined number of randomly selected pixel locations assigned a value of 1 and the remainder of pixels assigned a value of 0. A symmetric 2D Gaussian filter ($\sigma = 7$ pixels) was applied to the image, and all resultant values above 0.1 were set to 1, whereas the remaining values were set to 0. This produced a binary image with a number of randomly located contiguous regions with a value of 1. A second Gaussian blur ($\sigma = 1$ pixel) was applied to smooth the edges between 0- and 1-valued regions, producing the final 2D filter. The number of pixels originally assigned a value of 1 (i.e., prior to any blurring) corresponds to the number of “bubbles” in the filter. The filter was then multiplied with the modulation spectrum, effectively removing randomly selected sections of modulation energy from the signal and maintaining energy in the regions of the “bubbles.” Unlike in our previous study (Venezia et al., 2016), STMs less than 1 Hz and 0.5 cyc/kHz were always allowed to pass through the filter; this was found informally to reduce a “buzzing” distortion produced during signal processing. A filtered speech waveform was obtained from the degraded modulation spectrum by performing an inverse 2D Fourier transform followed by iterative spectrogram inversion (Griffin & Lim, 1984). For each of the 452 sentences, filtered versions were created

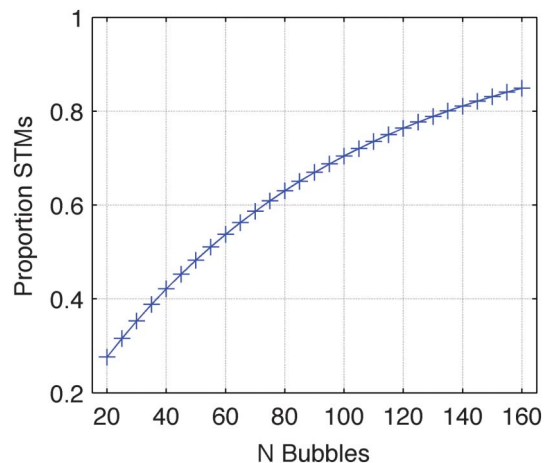
using independent random filter patterns; specifically, every sentence was processed with randomly selected patterns of bubbles ranging in number from 20 (little spectrotemporal information) to 160 (substantial spectrotemporal information) in steps of five bubbles (see Supplemental Material S1 for example stimuli). Thus, the total number of experimental stimuli was 13,108 (452 sentences \times 29 bubble steps). The relation between the number of bubbles and the average proportion of spectrotemporal information retained in the stimulus is shown in Figure 3.

All stimuli were generated offline and stored prior to the experiment. During the experiment, speech signals were delivered monaurally via a 24-bit soundcard (Envy24 PCI audio controller, VIA Technologies, Inc., or UA-101 USB audio interface, Edirol, Inc.), passed through a headphone buffer (HB6 or HB7, Tucker-Davis Technologies, Inc.), and presented to the listener through a Sennheiser HD600 headset. Stimuli were presented at a fixed overall level of 85 dB SPL, except for one of the YNH subjects for whom stimuli were presented at 75 dB SPL.

Procedure

All participants completed a pure-tone, air-conduction audiogram to ensure hearing thresholds were within the specified range for participation in the study. At Site 2, additional audiologic testing was performed including otoscopy, tympanometry, ipsilateral 1-kHz acoustic reflexes, and bone conduction audiograms. All testing was performed in a double-walled, sound-attenuated booth. At Site 1, an experimenter accompanied the participant inside the booth, whereas at Site 2, the experimenter communicated with the participant from outside the booth via intercom. The experiment began with a set of verbal instructions from the experimenter informing the participant that

Figure 3. Relation between number of bubbles (abscissa) and proportion of spectrotemporal information retained in the stimulus (ordinate). The actual bubbles step levels used in the experiment (20–160 in steps of five) are marked with crosses. STMs = spectrotemporal modulations.



acoustically distorted sentences would be played one at a time over headphones to the better-hearing ear. Participants were instructed to repeat the sentence back as completely as possible, including partial responses and guesses. On each trial, a single “bubble-ized” sentence was presented and the listener repeated the sentence he or she heard to the experimenter. A mouse and a keyboard were used by the experimenter to input and score responses. The experimenter typed the response into an edit box on the screen and denoted the keywords that were correctly identified in the response by clicking any of five corresponding buttons on the screen. Errors in tense were counted as correct. The next trial began after the scoring process was complete. An up–down staircase procedure was used to adjust the number of bubbles from trial to trial; if the listener correctly identified three or more keywords, the number of bubbles decreased by five, and if fewer than three words were correctly identified, the number of bubbles increased by five. This tracking procedure converged on a performance level in which listeners correctly identified three or more keywords on 50% of trials. The staircase procedure began at an initial step of 160 bubbles for the OHI listeners, 120 bubbles for the ONH listeners, and 100 bubbles for the YNH listeners. On each trial, the sentence was drawn pseudorandomly from the list of 452 possible sentences, and the version of that sentence with the current number of bubbles was presented to the listener. The same sentence was never repeated to a given listener. OHI listeners completed all trials without the assistance of hearing aids. The bubbles filter and listener response (zero to five keywords correct) from each trial were stored for later analysis. Listeners completed 16–18 blocks of 25 trials over two approximately 90-min sessions. Breaks were given between blocks as needed or at least every 30 min.

Analysis

CIImgs

The following subsections will provide a summary description of how CIImgs were derived and more thorough descriptions of how CIImgs were analyzed in individual listeners (first-level analysis) and compared across listener groups (second-level analysis). See Figure 4 for an overall schematic of the CIImg analysis.

Toward CIImgs: Ordinal regression analysis. The main purpose of this study was to determine the contribution of different STMs to intelligibility for each participant. This was achieved by identifying regions of the 2D STM spectrum (see Figure 1) whose inclusion in the stimulus systematically predicted good performance and whose removal systematically predicted poor performance. On each trial, a filter mask placed over the STM spectrum was “pierced” in randomly selected regions creating “bubbles” through which parts of the speech stimulus could be heard. If a region of the STM spectrum was important for sentence recognition, and if that region was revealed by “bubbles,” the participant would correctly identify many or all keywords. On the other hand, if this “important” region was

not revealed by bubbles, the participant would identify few or no keywords. This logic indicates that a regression between whether a region in the STM domain was or was not revealed (independent variable) and the number of keywords correct (dependent variable) should reveal the importance of that region for speech intelligibility. This is the essence of the formation of the CIImg, an ordinal regression between energy in different STM regions and the number of keywords correctly reported. The result is a set of weights describing the relative importance of each individual pixel in the 2D STM spectrum. The Appendix describes the details of how this regression was performed and evaluated given the large number of predictor variables (i.e., pixels) in the STM domain.

First-level analysis: CIImgs and model predictions for individual listeners. Each participant’s data were split into training and test sets and preprocessed as described in the Appendix, and a CIImg was derived. The CIImg weights were used to generate a vector of predicted responses for the training data, \hat{y}_{train} , and the test data, \hat{y}_{test} . The vector \hat{y}_{test} was compared to the true vector of responses in the test set to obtain an independent estimate of the regression error rate.

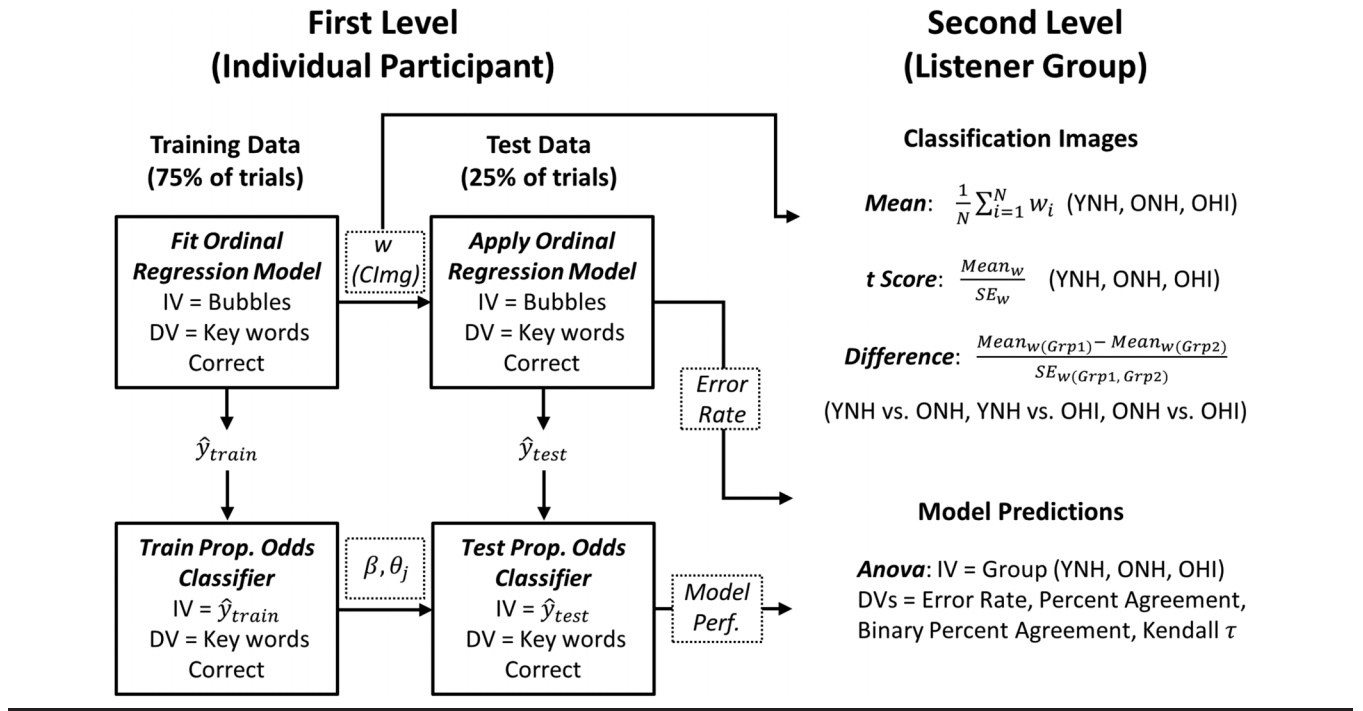
Because the ordinal regression procedure described in the Appendix generates predicted responses (\hat{y}) on a continuous scale, the model predictions cannot be used to estimate categorical accuracy, as is commonly done with ordinal classifiers. Therefore, the vector \hat{y}_{train} was treated as the input to a “proportional odds” ordinal classifier. The proportional odds model (McCullagh, 1980) is an extension of standard logistic regression in which the logit transformation is applied to cumulative response probabilities, γ_{ij} , as follows:

$$\text{logit}(\gamma_{ij}) = \log \frac{P(Y_i \leq j)}{1 - P(Y_i \leq j)} = \theta_j - x_i^T \beta, j = 1, \dots, J - 1 \quad (1)$$

where Y_i is the i th true response, j is a given response category, x_i is the vector of predictor variables for the i th observation, β is the set of corresponding regression parameters, and θ_j are parameters that provide each cumulative logit with an intercept. The model is referred to as “proportional odds” because there is a single regression slope, β , which applies equally to all response categories j . We used \hat{y}_{train} to train a proportional odds classifier to predict the true responses in the training data, and we then obtained independent response predictions from \hat{y}_{test} (output of the CIImg model applied to the test data). These predictions were compared with the true responses in the test data to obtain the following metrics of model performance: overall percent agreement, binary percent agreement (collapsing response categories to “incorrect,” zero to two keywords recognized, and “correct,” three to five keywords recognized), and Kendall’s τ (a measure of categorical correlation).

Second-level analysis: Comparison of CFImgs across listener groups. The primary goal of the CIImg analysis was

Figure 4. Classification image analysis schematic. The goal of first-level analysis was to generate for each participant a Clmg (w) using ordinal regression and to use the resultant Clmg model to predict the participant’s trial-by-trial responses (keywords correct) via a proportional odds classifier. Models were trained and tested on independent data. The Clmg (w) and model performance metrics (error rate, percent agreement, binary percent agreement, and Kendall τ) were passed to the second level for further analysis. The goal of second-level analysis was to calculate a mean and t -score Clmg for each listener group (YNH, ONH, and OHI) and to test for differences in the group level Clmg statistically (difference t score). The listener group means of Clmg model performance metrics were also compared statistically at the second level. IV = independent/predictor variable; DV = dependent/criterion variable; Clmg = classification image; SE = standard error; YNH = young normal-hearing; OHI = older hearing-impaired; ONH = older normal-hearing.



to test for systematic differences in the image weights across the three listener groups. To accomplish this, a mean Clmg was constructed for each group and these mean images were compared statistically across groups. Specifically, the first-level Clmgs obtained from each participant were smoothed using a Gaussian blur ($\sigma = 5$ pixels) and averaged across participants within a listener group, yielding mean Clmgs for the YNH, ONH, and OHI groups. These mean Clmgs were transformed to t -statistic images separately for each listener group by dividing the mean at each pixel by the standard error of the mean across participants. “Difference t -statistic” images were calculated for each pair of listener groups (YNH vs. OHI, ONH vs. OHI, and YNH vs. ONH) by differencing the mean Clmgs and scaling the resultant image by the unpooled standard error. “Difference p -value” images were obtained from the difference t -statistic images. These p -value images were corrected for multiple comparisons using the false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995), such that pixel-wise differences in the Clmgs were considered significant if the FDR-adjusted $p < .05$.

Clmg model predictions obtained at the first level were also compared across listener groups at the second level. Percent agreement, binary percent agreement, and

Kendall’s τ , along with the independent estimate of the ordinal regression error rate, were analyzed for group mean differences using Welch’s analysis of variance (ANOVA) and the Games–Howell post hoc testing procedure, which are robust to unequal sample size and variance (Field, 2013).

Threshold Performance

For each participant, an up–down staircase was implemented in which the number of bubbles (i.e., amount of STM information in the speech signal) was adjusted such that the listener identified three or more keywords (i.e., got the trial “correct”) on ~50% of trials. Therefore, performance on the task can be quantified in terms of the number of bubbles required to achieve 50% correct performance. This was estimated by averaging the number of bubbles across all trials in the experiment excluding the first block of 25 trials. The average number of bubbles was converted to a more direct measure of the proportion of pixels in the modulation spectrum revealed to the listener. This was done by calculating for each number of bubbles in the stimulus set (20–160 in steps of five) the average proportion of pixels with a value of “1”—that is, those pixels allowing modulation information to pass through the bubbles filter. The relationship between number of bubbles

and proportion of pixels revealed took an exponential form, so we fit a two-parameter exponential function as follows:

$$Prop_{STM} = 1 - \alpha e^{-\beta \cdot N_{bub}}, \quad (2)$$

where $Prop_{STM}$ is the proportion of STM pixels revealed, N_{bub} is the number of bubbles, and the fitted parameters are α and β . The best-fitting values of α and β were 0.906 and 0.011, respectively (see Figure 3). $Prop_{STM}$ was then estimated for each listener by entering that listener's average number of bubbles into the fitted equation as N_{bub} .

Listener group differences in $Prop_{STM}$ were tested using Welch's ANOVA and the Games–Howell post hoc testing procedure. The effects of age, hearing status (high-frequency average [HFA] at 2–6 kHz), and two CImg parameters (described in the Results section below) on $Prop_{STM}$ were examined separately in the ONH and OHI groups. Specifically, a bidirectional stepwise regression was performed to select the best-fitting model as assessed by the Akaike information criterion (Akaike, 1974).

Results

Outline

First-level (individual participant) CImgs were calculated to determine a set of weights in the STM domain that best predicted the trial-by-trial responses given by each participant. Second-level (listener-group average) CImgs were then calculated to determine whether these weights differed across listener groups. CImg model predictive performance was also compared across listener groups at the second level. Additionally, threshold performance in the behavioral task ($Prop_{STM}$) was compared across listener groups and within the older listener groups, and assessed for a significant relationship with age, hearing status, and CImg parameters.

CImgs

Main Analyses

Second-level CImgs. The second-level mean and t -score CImgs for the YNH, ONH, and OHI listener groups are plotted in Figure 5. For the mean CImgs (top row), warm colors indicate relatively strong contributions to intelligibility, whereas cool colors indicate little or no contribution to intelligibility. For the t -score CImgs (bottom row), colors index the magnitude of the contribution to intelligibility relative to the variability in this magnitude across participants. From the mean CImg, it is immediately clear that there is very little, if any, difference in the pattern of weights across the listener groups. For all three groups, the primary contributions to intelligibility come from a “hot spot” centered between 1–7 Hz and 0–1.5 cyc/kHz. This is in agreement with our previous work (Venezia et al., 2016). The t -score images suggest that the effects were less reliable (smaller t scores near the CImg “hot spot”) in the OHI

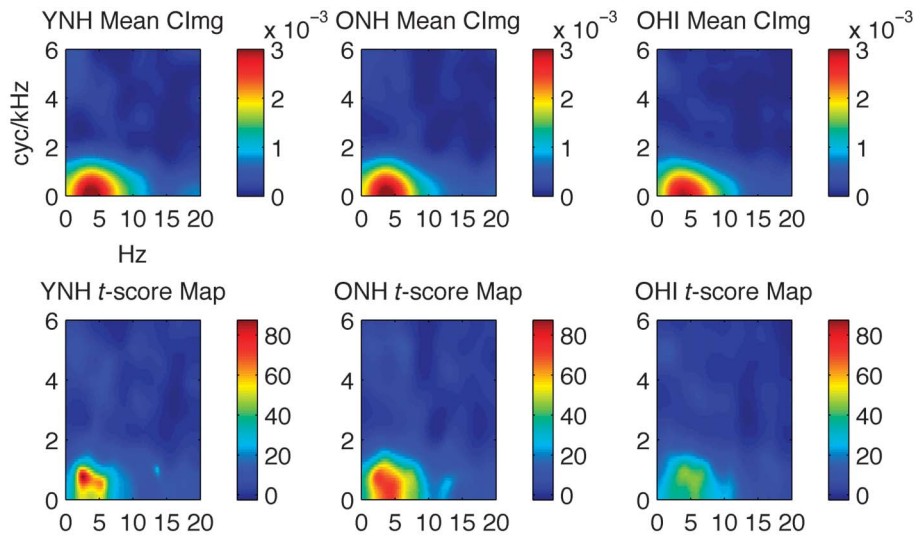
group. Direct comparison of the mean CImgs across the different pairwise combinations of listener groups failed to reveal any significant differences at the FDR-corrected $p < .05$ level. However, CImg weights near the upper edge of the “hot spot” were observed to be larger in the YNH and ONH groups compared to the OHI group at an uncorrected threshold of $p < .001$ (not pictured). Together with the smaller CImg t scores (higher variance) in the OHI group (see Figure 5, bottom row), this suggests that a systematic effect of hearing loss on CImg weights was not detected in the group comparisons due to the large range of hearing losses in the OHI group. Indeed, there was some overlap of thresholds in the OHI/ONH groups toward the lower end of the OHI range (see Figure 2, purple). An explicit test of the relation between hearing thresholds and CImg weights is carried out in the post hoc analyses below (see Post Hoc Analyses section).

Predictions of trial-by-trial responses. To determine whether the individual participant CImgs were equally capable of predicting trial-by-trial listener behavior across the three listener groups, we examined group differences in CImg model predictions. Table 1 summarizes CImg model predictions in terms of four performance metrics—regression error rate, which was estimated directly from the ordinal regression model used to define the CImg (see Toward CImgs: Ordinal Regression Analysis section and the Appendix); percent agreement; binary percent agreement; and Kendall's τ —which were estimated from the proportional odds classifier (see First-Level Analysis: CImgs and Model Predictions for Individual Listeners section). All metrics were calculated from independent “test” data that were not used to train the model. The table indicates a consistent pattern: (a) The predictive power of the model is good across all metrics; (b) significant listener-group differences were observed for all metrics (ANOVA column); (c) the best predictions were observed for the YNH group, followed by the ONH group and the OHI group (mean columns); and (d) pairwise differences were only significant when comparing the normal-hearing groups to the OHI group (comparison columns). Overall, poorer model performance in the OHI group indicates that factors beyond the STM patterns in the stimuli accounted for a comparatively larger share of the variance in trial-by-trial performance.

Post Hoc Analyses

The main analyses (a) suggested an effect of hearing loss on CImg weights and (b) demonstrated that CImg model predictions are significantly less accurate for OHI listeners. Differences in CImg weights likely reflect a systematic effect of hearing loss—that is, as the degree of hearing loss increases, the magnitude of the weights systematically declines in a particular region of the CImg. On the other hand, poor CImg model predictive performance could arise from a variety of factors (e.g., reduced audibility, suprathreshold distortion, changes in cognition, or a combination of these factors) and thus may be reflected in the CImg weights as nonsystematic changes or “noise” that

Figure 5. Second-level Clmgs. Top row: mean Clmgs; color bar indicates ordinal regression weights. Bottom row: one-sample *t*-score images; color bar indicates *t*-score magnitude. Listener groups are labeled in the titles above each panel. Axis labels on the top left panel apply to all panels. Ordinate: spectral modulation rate (cyc/kHz). Abscissa: temporal modulation rate (Hz). Clmg = classification image; YNH = young normal-hearing; OHI = older hearing-impaired; ONH = older normal-hearing.



scales with degree of hearing loss. We carried out two post hoc analyses to test these possibilities.

To test for systematic changes in Clmg weights due to hearing loss, we calculated the pixel-wise correlation between HFA thresholds and Clmg weights across OHI listeners to form a “correlation Clmg.” This image was converted to a *t*-score image and tested for significance using the FDR procedure. Figure 6 shows that, indeed, the magnitudes of Clmg weights near the upper edge of the “hot spot” (i.e., where trend-level group differences were observed) were negatively correlated with HFA. However, caution must be exercised in attributing this effect solely to hearing status because HFA was correlated with age in the OHI group

($r = .57$). To examine the effect of age, we constructed correlation Clmgs for age in the ONH and OHI listeners. No significant correlations with age were observed for either of the older listener groups, suggesting that hearing loss, not age, is responsible for the systematic shift in the Clmg weights in the region shown in Figure 6 for OHI listeners. We will henceforth refer to the average magnitude of the weights in this region using D-SYS to reflect the degree of systematic distortion of the Clmg in a given OHI listener (e.g., see Figure 6, right panel, ordinate). Note that lower scores on D-SYS indicate greater Clmg distortion.

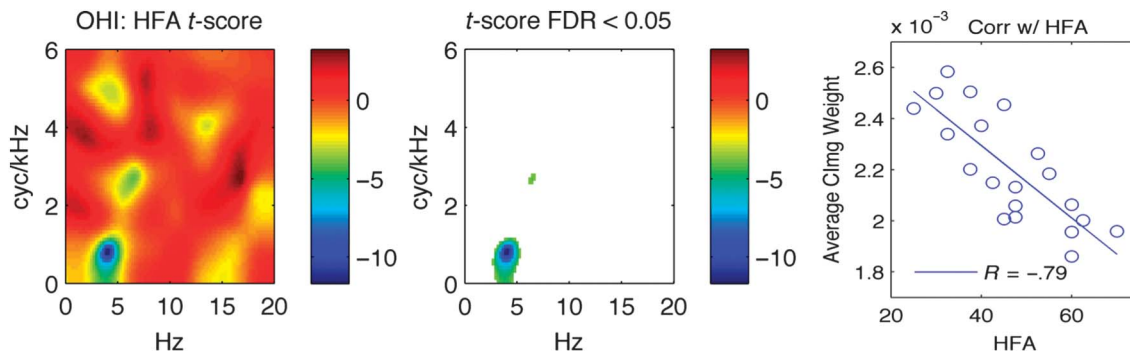
To test for nonsystematic changes in Clmg weights due to hearing loss, we calculated a separability index for

Table 1. Group means and statistics for classification-image model performance metrics.

	Mean (95% CI)			Welch's ANOVA			YNH vs. OHI		ONH vs. OHI		YNH vs. ONH	
	YNH	ONH	OHI	<i>F</i> (<i>df</i>)	<i>p</i>	Adj ω^2	<i>t</i> (<i>df</i>)	<i>p</i>	<i>t</i> (<i>df</i>)	<i>p</i>	<i>t</i> (<i>df</i>)	<i>p</i>
Regression error	0.11 (0.014)	0.12 (0.012)	0.16 (0.022)	8.5 (2, 27.5)	.001	0.23	-4.2 (27.6)	.002	-3.4 (28.8)	.012	-1.3 (21.9)	.423
Percent agreement	60.5 (4.26)	58.4 (3.35)	53.0 (4.22)	3.8 (2, 26.7)	.034	0.1	2.7 (25.6)	.030	2.1 (35.5)	.106	0.86 (21.1)	.670
Binary percent agreement	86.3 (1.98)	85.5 (1.27)	82.6 (2.51)	5.3 (2, 25.0)	.011	0.15	3.2 (27.9)	.010	2.9 (28.0)	.018	0.73 (17.6)	.751
Kendall τ	0.72 (0.038)	0.71 (0.021)	0.64 (0.040)	5.9 (2, 23.8)	.008	0.17	3.2 (26.4)	.027	3.2 (28.9)	.027	0.56 (15.8)	.840

Note. Regression error is the error rate in the ordinal regression procedure used to define first-level classification image weights; percent agreement is the percentage of trials for which the first-level proportional odds classifier predicted the correct number of keywords recognized (0–5); binary percent agreement is calculated in the same way as percent agreement, but collapsing responses into two categories (“correct” = more than two keywords recognized, “incorrect” = two or fewer keywords recognized). The width of the 95% confidence interval (CI) is displayed for the group means. Effect size (adjusted omega squared) is displayed for Welch’s ANOVA tests. Post hoc *t* tests have been adjusted using the Games–Howell approach. Significant *p* values (< .05) are shown in bold. ANOVA = analysis of variance; YNH = young normal-hearing; OHI = older hearing-impaired; ONH = older normal-hearing; Adj = adjusted.

Figure 6. Correlation of high-frequency average pure-tone thresholds with CImg weights in OHI listeners. Left: pixel-wise correlation of HFA with CImg weights in the modulation spectrum domain (abscissa: Hz, ordinate: cyc/kHz), expressed as t scores. Middle: same as left but nonsignificant pixels (FDR-adjusted $p > .05$) are displayed as white background. Right: HFA (abscissa) is plotted against the average classification weight (ordinate) taken across the significant pixels identified in the middle panel. The best-fitting regression line and magnitude of the correlation are shown. CImg = classification image; FDR = false discovery rate; HFA = high-frequency average; YNH = young normal-hearing; OHI = older hearing-impaired; ONH = older normal-hearing.



the CImg of each individual listener. Separability refers to the extent to which a 2D CImg can be reconstructed from separate, one-dimensional response functions in the spectral and temporal domains without any loss of information (Depireux, Simon, Klein, & Shamma, 2001). An index of separability was calculated for the CImg of each listener using singular value decomposition (Depireux et al., 2001). This separability index ranged from 0 (*completely nonseparable*) to 1 (*completely separable*). Bubbles-based CImgs for intelligibility tend to be highly separable—indeed, the separability index for the mean CImg in our original sample of YNH listeners (Venezia et al., 2016) was greater than 0.99. Thus, we hypothesized that deviations from separability (i.e., lower scores on the index) could be used to capture nonsystematic deviations of CImg weights from the typical pattern. We will henceforth refer to this measure as D-NOISE to reflect the degree of nonsystematic distortion of the CImg in a given listener. Note that lower scores on D-NOISE indicate greater CImg distortion. In Figure 7, we plot the distribution of D-NOISE across YNH, ONH, and OHI listeners along with examples of the individual participant CImgs with the highest and lowest D-NOISE. There was a significant effect of listener group on D-NOISE, Welch's $F(2, 29.3) = 10.3, p < .001$, est. $\omega^2 = 0.28$, with significantly lower values in OHI listeners compared to YNH, Games–Howell $t(25.5) = 4.6, p < .001$, and ONH, Games–Howell $t(27.7) = 3.5, p < .01$, listeners.

Thus, D-NOISE patterned in the same way across listener groups as CImg model predictive performance (OHI significantly different from YNH/ONH). A multiple regression across OHI listeners with D-SYS and D-NOISE as predictor variables and Kendall's τ (CImg model accuracy) as the criterion variable showed that only D-NOISE was significantly associated with CImg model predictive performance, $b = 0.75, t(17) = 4.1, p < .001$. This suggests that D-NOISE/CImg model performance reflects a processing factor (or factors) that causes performance in

OHI listeners to be less strongly coupled to stimulus information alone. An ideal observer simulation showed that changes in D-NOISE could be induced by varying the degree of a virtual observer's internal noise but not by changing the number of bubbles in the speech stimuli (see Supplemental Material S2). Interestingly, we show in the subsequent section that D-NOISE also predicts unique variance in threshold performance after accounting for differences in hearing thresholds.

Threshold Performance

Proportion of STMs Revealed

The average proportion of spectrotemporal information revealed to the listener at the 50% correct performance level, $Prop_{STM}$, differed significantly between listener groups, Welch's $F(2, 29.8) = 45.2, p < .001$, est. $\omega^2 = 0.64$. Games–Howell-adjusted post hoc tests revealed that $Prop_{STM}$ was significantly smaller for YNH than OHI (diff = 0.20), $t(24.1) = 8.6, p < .001$, and ONH (diff = 0.08), $t(26.2) = 6.1, p < .001$, and smaller for ONH than OHI (diff = 0.12), $t(26.8) = 5.0, p < .001$. That is, YNH listeners performed best on the task because they were able to obtain 50% correct with the smallest amount of spectrotemporal information, followed by ONH and then OHI (see Figure 8, left panel).

A stepwise regression was performed to determine which factors could best account for the pattern of performance in the older listener groups. For the ONH group, the candidate predictors were HFA and age. For OHI listeners, the candidate predictors were HFA, age, D-SYS, and D-NOISE. For the ONH listeners, the most parsimonious model included only HFA as a predictor, but HFA was not significantly correlated with $Prop_{STM}$ ($r = .32, p = .180$), suggesting that neither age nor HFA can explain differences in performance across the ONH listeners. For the OHI group, the most parsimonious model included all four predictors, but only the regression coefficients for

Figure 7. Nonsystematic classification image (Clmg) distortion (D-NOISE) across listener groups (YNH, ONH, and OHI). Left: distribution plots of D-NOISE (lower = more Clmg distortion) for each listener group. White horizontal bars show the group mean, gray shaded regions show the 95% confidence interval of the mean, and colored regions show the standard deviation. Cyan markers show the individual participant scores. Statistically significant differences between groups are marked by black bars. Middle: Clmg for the single listener with the highest D-NOISE (least Clmg distortion). Right: Clmg for the single listener with the lowest D-NOISE (greatest Clmg distortion). YNH = young normal-hearing; OHI = older hearing-impaired; ONH = older normal-hearing.

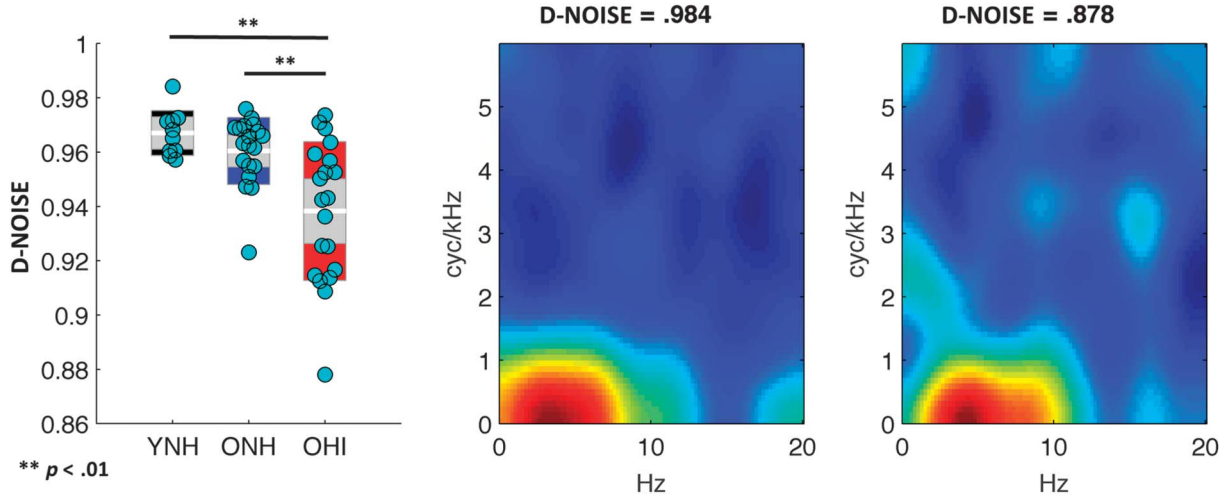
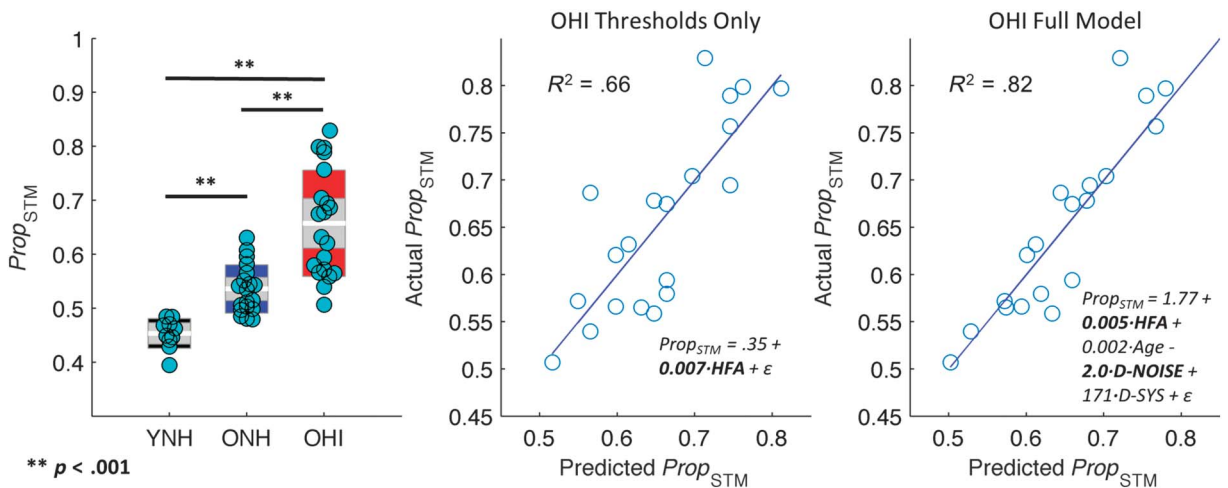


Figure 8. Threshold behavioral performance ($Prop_{STM}$, smaller = better performance) and relation to hearing status and classification image (Clmg) model parameters in OHI listeners. Left: distribution plots showing the across-trial average proportion of STMs revealed to listeners in each group (YNH, ONH, and OHI). White horizontal bars show the group mean, gray shaded regions show the 95% confidence interval of the mean, and colored regions show the standard deviation. Cyan markers show the individual participant scores. Statistically significant differences between groups are marked by black bars. Middle: scatter plot of predicted threshold performance (abscissa) against actual threshold performance (ordinate) in OHI listeners for a model including high-frequency average (HFA) thresholds as a single predictor. Regression equation is shown along with the best-fitting regression line and model R^2 . Significant predictors ($p < .05$) are bolded in the regression equation. Right: scatter plot of predicted threshold performance (abscissa) against actual threshold performance (ordinate) in OHI listeners for a model including HFA thresholds, age, Clmg separability (D-NOISE), and the average magnitude of Clmg weights in the “hot spot” (D-SYS). Regression equation is shown along with the best-fitting regression line and model R^2 . Significant predictors ($p < .05$) are bolded in the regression equation. Model fit is improved when Clmg parameters are included. YNH = young normal-hearing; OHI = older hearing-impaired; ONH = older normal-hearing; STM = spectrotemporal modulation.



HFA, $b = 0.66$, $t(15) = 3.5$, $p < .01$, and D-NOISE, $b = -0.53$, $t(15) = 3.4$, $p < .01$, were significantly different from zero. HFA uniquely accounted for 14% of the variance in $Prop_{STM}$, whereas D-NOISE uniquely accounted for 13%, and common variance among the predictors accounted for 47%. This demonstrates that D-NOISE accounts for additional variance in threshold performance over and above the variance accounted for by hearing thresholds. Figure 8 (center, right panels) compares the regression model predictions of an HFA-only model versus the full model selected by stepwise regression in OHI listeners. A significantly better fit is obtained for the full model, that is, when CImg parameters are considered. This suggests that D-NOISE reflects one or more aspects of suprathreshold auditory and/or cognitive processing in OHI listeners.

Discussion

The primary aim of this study was to determine whether age and/or hearing status affects the relative importance of different regions of the speech STM spectrum for intelligibility. To accomplish this, we used our recently devised psychoacoustic procedure, auditory bubbles (Venezia et al., 2016), to determine for YNH, ONH, and OHI listeners a set of weights—a CImg—showing which regions of the speech modulation spectrum contribute most to intelligibility. We used an ordinal regression procedure to determine the best-fitting set of weights (Measure 1) and, using independent data, measured how accurately the resulting regression model predicted trial-by-trial performance in individual listeners (Measure 2).

As noted in the introduction of this study, changes in Measures 1 and/or 2 could feasibly be produced by a loss of audibility, an increase in peripheral or central suprathreshold distortion, a change in cognition/listening strategy, or some combination thereof. The auditory bubbles method cannot isolate effects produced by any one of these sources. To mitigate differences in audibility, we presented speech at a level (85 dB SPL overall) that exceeded hearing thresholds at low to mid frequencies in all listeners. We also included hearing thresholds as covariates in our statistical models. To minimize cognitive demands, we asked listeners to perform speech recognition in quiet rather than in noise. These steps were intended to allow us to relatively amplify the effects of peripheral or central suprathreshold distortion on Measures 1 and 2.

In fact, the patterns of CImg weights (Measure 1) did not vary strongly across the listener groups, but weights in the region of primary importance for intelligibility appeared to be more variable in OHI group. At a more lenient statistical threshold, it was revealed that CImg weights tended to be reduced for OHI listeners in this region. A post hoc analysis confirmed a strong negative correlation between CImg weights and HFA thresholds in this region. Thus, hearing loss appears to produce a systematic distortion (D-SYS) of CImgs in the region of primary importance to intelligibility. Additionally, CImg model predictions were significantly worse for OHI listeners compared to YNH/ONH

listeners. We hypothesized that this could be related to an increase in internal noise, which may reflect deficits at multiple levels of processing. In fact, the degree of non-systematic CImg distortion (D-NOISE) was found to be highly predictive of CImg model predictive performance in OHI listeners, and a simulation demonstrated that increasing the internal noise of a “virtual observer” produced systematic changes in D-NOISE.

A secondary goal was to examine listener group differences in threshold performance. The amount of STM information removed from the signal was varied adaptively from trial to trial; thus, we measured the amount of spectrotemporal distortion a listener was able to tolerate—that is, the average proportion of STMs retained in the signal ($Prop_{STM}$)—to achieve a performance level of 50% correct. This measure reflects the “distortion sensitivity” of a listener (Houtgast & Festen, 2008), which has long been attributed to suprathreshold factors. We compared this quantity across listener groups and examined the effects of age, hearing status, D-SYS, and D-NOISE on threshold performance. Threshold performance was significantly better in YNH listeners followed by ONH and then OHI listeners. Poor threshold performance in OHI listeners was significantly associated with HFA and D-NOISE, which, although positively correlated, each accounted for ~ 15% of unique variance in performance. The following sections review our main findings and their significance.

ONH and OHI Listeners Rely on the Same STMs as YNH Listeners

The clearest evidence of suprathreshold processing effects in ONH and/or OHI listeners would have been a shift in the large-scale pattern of CImg weights compared to YNH listeners. However, our data did not bear this out. In fact, there were no significant differences in the CImg weights between listener groups (see Main Analyses section). Moreover, when we examined the CImgs of individual listeners, it was clear that listeners relied on a well-defined region of the speech modulation spectrum (< 10 Hz and < 2 cyc/kHz with a peak at ~4 Hz in the temporal modulation domain) regardless of age or hearing status. Thus, the potential outcome discussed in the introduction in which OHI/ONH listeners would rely relatively more on lower spectral and temporal modulation rates than YNH listeners was not supported.

However, our negative finding does not necessarily run counter to the existing evidence. Our work and those of others show that even YNH listeners tend to rely primarily on relatively low-rate STMs for speech intelligibility (Chi et al., 1999; Drullman et al., 1994a, 1994b; Elliott & Theunissen, 2009; Ter Keurs et al., 1992, 1993a; Venezia et al., 2016). Spectral and temporal processing deficits associated with aging and/or hearing loss may not emerge at such low rates or even at higher rate STMs toward the edges of the CImg (Bacon & Viemeister, 1985; Florentine, Fastl, & Buus, 1988; Gifford, Bacon, & Williams, 2007; Henry et al., 2005; Snell, 1997; Strouse, Ashmead, Ohde, &

Grantham, 1998; Turner, Souza, & Forget, 1995; Walton, Orlando, & Burkard, 1999). Moreover, temporal and spectral modulation transfer functions measured in older and/or hearing-impaired listeners (Bacon & Viemeister, 1985; Grant et al., 1998; Shen, 2014; Ter Keurs et al., 1993b) suggest that suprathreshold deficits reflect changes in sensitivity (an overall shift in performance across modulation rates) rather than cutoff frequency (a change in the pattern of performance across modulation rates). Therefore, OHI listeners may process the low-rate STMs in speech less efficiently than YNH listeners. This is consistent with the findings of Bernstein and colleagues (Bernstein, 2016; Bernstein et al., 2013, 2016; Mehraei et al., 2014). In the following section, we review results from the current study that further support this conclusion for hearing-impaired listeners in particular.

OHI Listeners Process Crucial STMs Less Efficiently

Although no large-scale differences were observed in the pattern of CImg weights across listener groups, there was evidence of a systematic decrease in CImg weights in the region of primary importance to intelligibility. This size of decrease (D-SYS) scaled with the degree of high-frequency hearing loss (HFA, average pure-tone threshold from 2 to 6 kHz) in OHI listeners. Specifically, D-SYS was negatively correlated with HFA ($r = -.79$) in the OHI listener group (see Figure 6). This resulted in increased variability and thus lower CImg t scores for OHI listeners compared to ONH and YNH listeners (see Figure 5). Additionally, CImg regression models were less able to predict the trial-by-trial responses of OHI listeners compared to ONH or YNH listeners (see Table 1). We therefore suggest that, although all of the listener groups relied on the same STMs for intelligibility, OHI listeners used this information less efficiently; that is, although the trial-by-trial STM patterns conveyed to the listener predicted much of the variance in trial-by-trial performance in all three listener groups (see Table 1), other factors contributed relatively more to OHI performance.

What is the nature of these other factors? A straightforward conclusion is that reduced audibility contributed to unexplained trial-by-trial variance in OHI listeners. This is consistent with the strong negative correlation observed between D-SYS and HFA and with the fact that significant differences in CImg model predictions were only observed between the OHI group and the normal-hearing groups, but not between ONH and YNH listeners. However, when we pitted D-SYS against D-NOISE—a measure of nonsystematic CImg distortion—to test which CImg parameter was most strongly associated with model predictive accuracy in OHI listeners, a significant association was observed only for D-NOISE. In addition, although D-NOISE was also strongly correlated with HFA, it predicted unique variance in threshold performance across listeners in the OHI group (i.e., after accounting for variance due to HFA; see Figure 8). Simulations showed that an increase in internal

noise produces a roughly linear decrease in D-NOISE (an increase in CImg distortion). An increase in internal noise could, in theory, result from reduced audibility, suprathreshold distortion, or impaired cognition, but we can partially rule out audibility because D-NOISE is independently associated with speech recognition performance after accounting for HFA. It is also tempting to rule out cognition because the ONH listeners—who should be cognitively impaired relative to YNH listeners (see introductory paragraphs)—did not show increased D-NOISE compared to YNH listeners. However, we should note that deficits in cognition interact (i.e., become worse) with deficits in hearing loss (Lin et al., 2013), and these compounded deficits may have produced the noted shift in D-NOISE in OHI listeners. Our data therefore provide motivation to investigate suprathreshold auditory and cognitive factors as sources of individual differences in D-NOISE. These may include deficits in sustained attention or executive function (Chao & Knight, 1997) or deficits in the integration of STM information across frequency channels (Healy & Bacon, 2002; Souza & Boike, 2006; Turner, Chi, & Flock, 1999), neither of which would be expected to produce systematic changes in CImgs.

Young Listeners Outperform OHI Listeners

YNH listeners required significantly less STM information than ONH and OHI listeners to achieve 50% correct performance on our sentence recognition task. Additionally, ONH required less information than (i.e., outperformed) OHI listeners. Reduced audibility combined with suprathreshold auditory and/or cognitive processing deficits in OHI listeners (see discussion of D-NOISE above) likely accounted for the latter difference. However, audibility likely did not account for differences between younger and older listeners. Specifically, a model of speech intelligibility (the Coherence Speech Intelligibility Index; Kates & Arehart, 2005) predicted better performance in the YNH listeners even after adjusting the performance of ONH and OHI listeners for poorer audibility (see Supplemental Material S2). Therefore, threshold performance on the auditory bubbles task appears to be sensitive to factors beyond audibility.

Which of these factors is likely to account for the difference in performance between the YNH group and the two older listener groups? There was no significant correlation between age and $Prop_{STM}$ in the ONH group, and performance in the OHI group was better accounted for in terms of HFA and D-NOISE compared to age (see Proportion of STMs Revealed section). Therefore, effects of age were observed between the listener groups but not within the older groups, as might be expected. One possible reason for this is that we did not include many middle-aged listeners in the older groups. More than 80% of our older listeners were over the age of 60 years. However, recent research indicates that age-related temporal processing and cognitive decline—each of which may have influenced performance on the task—begin to emerge in middle age (Grose, Mamo, Buss, & Hall, 2015; Gunstad et al., 2006;

Helfer, 2015). Therefore, gradations in performance may be more noticeable between middle-aged and older listeners, rather than among older listeners alone, as was the case with younger and older listeners here. Indeed, the group-level age effect we observed was rather small to begin with, so it may be the case that many more participants would be required to detect a correlation with age in a cohort of older listeners, especially because the magnitude of age-related decline varies widely across individuals (Gunstad et al., 2006). Another interesting question concerns whether the group-level age effect we observed is more likely to be explained by a decline in cognition versus a deficit in suprathreshold (likely temporal) auditory processing. Because we did not measure cognitive ability or temporal processing explicitly, we cannot provide a definitive answer. However, we can speculate that cognition played at least some role. Work by Wingfield (1996) suggests that, although older listeners generally receive a larger benefit from contextual information than young listeners on difficult speech recognition tasks, older listeners are less able to use this information when the context is presented after an ambiguous section of speech. This is because working memory limits the ability for retrospective analysis in older listeners. Our data suggest that the older listeners were often faced with scenarios involving retrospective analysis—that is, of trials for which one to two keywords were correctly identified, at least one of those keywords was in the sentence-final or penultimate key word position more than 70% of the time (ONH: 75%, OHI: 72%). Thus, older listeners were perhaps robbed of contextual benefits that could have narrowed the rather small performance gap between themselves and the younger listeners. Additional work will be necessary to determine whether and to what extent cognition and suprathreshold auditory measures predict performance on the auditory bubbles task.

Summary

We hypothesized that hearing loss and/or normal aging would produce changes in the perceptual representation of speech STMs. This was tested by constructing CImgs showing which STMs are crucial for intelligibility—essentially “perceptual receptive fields”—and comparing those images across groups of YNH, ONH, and OHI listeners. Listeners performed a sentence recognition task in which random patterns of STMs were retained in the stimulus from trial to trial. A regularized ordinal regression relating these patterns to listener responses (keywords correctly identified) was used to train CImg models, and we tested the ability of the models to predict individual listeners’ responses using independent data. We also recorded the proportion of STM information required to achieve 50% correct performance for each listener. Although we did not observe large-scale differences in patterns of the CImgs among the listener groups, OHI listeners exhibited both systematic and nonsystematic “distortions” in their patterns of CImg weights. The degree of CImg distortion was associated with the threshold performance of individual listeners

even after accounting for loss of audibility. Additionally, YNH listeners performed better overall than ONH or OHI listeners. We speculate that differences in cognition partially explain this effect, although changes in suprathreshold auditory function likely also contribute. Overall, we conclude that YNH, ONH, and OHI listeners rely on the same speech STM information for intelligibility, but these groups separate in terms of how efficiently this information is encoded (YNH > ONH > OHI).

Acknowledgments

This work was supported by an American Speech-Language-Hearing Foundation New Investigators Research Grant to J. H. V. and a UC Irvine Undergraduate Research Opportunities Program award to A.-G. M. Research reported in this publication was supported by the National Institute on Deafness and Other Communication Disorders under Award R21 DC013406 (Multiple principal investigators: V. M. R. and Y. Shen), the National Center for Research Resources and the National Center for Advancing Translational Sciences under Award UL1 TR001414, and the UC Irvine Alzheimer’s Disease Research Center under Award P50 AG016573, all from the National Institutes of Health. We thank Marjorie Leek and Yi Shen for constructive feedback on earlier versions of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This material is the result of work supported with resources and the use of facilities at the VA Loma Linda Healthcare System, Loma Linda, CA. The contents do not represent the views of the U.S. Department of Veterans Affairs or the U.S. Government.

References

- Akaike, H. (1974). A new look at the statistical model identification. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 215–222). New York, NY: Springer.
- Albinet, C. T., Boucard, G., Bouquet, C. A., & Audiffren, M. (2012). Processing speed and executive functions in cognitive aging: How to disentangle their mutual relationship. *Brain and Cognition, 79*, 1–11.
- Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2012). Aging affects neural precision of speech encoding. *The Journal of Neuroscience, 32*, 14156–14164.
- Arai, T., & Greenberg, S. (1997). The temporal properties of spoken Japanese are similar to those of English. Paper presented at EUROSPEECH, Rhodes, Greece.
- Bacon, S. P., & Viemeister, N. F. (1985). Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. *Audiology, 24*, 117–134.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological), 57*, 289–300.
- Bernstein, J. G. (2016). Spectrotemporal modulation sensitivity as a predictor of speech intelligibility in noise for hearing-impaired listeners. *The Journal of the Acoustical Society of America, 139*, 2121.
- Bernstein, J. G., Danielsson, H., Hällgren, M., Stenfelt, S., Rönnberg, J., & Lunner, T. (2016). Spectrotemporal modulation sensitivity as a predictor of speech-reception performance in noise with hearing aids. *Trends in Hearing, 20*, 2331216516670387.

- Bernstein, J. G., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., & Leek, M. R.** (2013). Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. *Journal of the American Academy of Audiology, 24*, 293–306.
- Braver, T. S., & West, R.** (2008). Working memory, executive control, and aging. *The Handbook of Aging and Cognition, 3*, 311–372.
- Cardin, V.** (2016). Effects of aging and adult-onset hearing loss on cortical auditory regions. *Frontiers in Neuroscience, 10*, 199.
- Caspary, D. M., Ling, L., Turner, J. G., & Hughes, L. F.** (2008). Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *Journal of Experimental Biology, 211*, 1781–1791.
- Chao, L. L., & Knight, R. T.** (1997). Prefrontal deficits in attention and inhibitory control with aging. *Cerebral Cortex, 7*, 63–69.
- Chi, T., Gao, Y., Guyton, M. C., Ru, P., & Shamma, S.** (1999). Spectro-temporal modulation transfer functions and speech intelligibility. *The Journal of the Acoustical Society of America, 106*, 2719–2732.
- Chi, T., Ru, P., & Shamma, S. A.** (2005). Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America, 118*, 887–906.
- Cohen, R. L., & Keith, R. W.** (1976). Use of low-pass noise in word recognition testing. *Journal of Speech and Hearing Research, 19*, 48–54.
- Cooper, J. C., & Cutts, B. P.** (1971). Speech discrimination in noise. *Journal of Speech and Hearing Research, 14*, 332–337.
- Depireux, D. A., Simon, J. Z., Klein, D. J., & Shamma, S. A.** (2001). Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *Journal of Neurophysiology, 85*, 1220–1234.
- Dirks, D. D., Morgan, D. E., & Dubno, J. R.** (1982). A procedure for quantifying the effects of noise on speech recognition. *Journal of Speech and Hearing Disorders, 47*, 114–123.
- Drullman, R., Festen, J. M., & Plomp, R.** (1994a). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America, 95*, 2670–2680.
- Drullman, R., Festen, J. M., & Plomp, R.** (1994b). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America, 95*, 1053–1064.
- Dubno, J. R., Dirks, D. D., & Morgan, D. E.** (1984). Effects of age and mild hearing loss on speech recognition in noise. *The Journal of the Acoustical Society of America, 76*, 87–96.
- Duquesnoy, A. J.** (1983). The intelligibility of sentences in quiet and in noise in aged listeners. *The Journal of the Acoustical Society of America, 74*, 1136–1144.
- Elhilali, M., Chi, T., & Shamma, S. A.** (2003). A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. *Speech Communication, 41*, 331–348.
- Elliott, T. M., & Theunissen, F. E.** (2009). The modulation transfer function for speech intelligibility. *PLoS Computational Biology, 5*, e1000302.
- Field, A.** (2013). *Discovering statistics using IBM SPSS statistics*. Thousand Oaks, CA: Sage.
- Fitzgibbons, P. J., & Gordon-Salant, S.** (2010). Behavioral studies with aging humans: Hearing sensitivity and psychoacoustics. In S. Gordon-Salant, R. Frisina, A. Popper, & R. Fay (Eds.), *The aging auditory system. Springer handbook of auditory research* (Vol. 34, pp. 111–134). New York, NY: Springer.
- Florentine, M., Fastl, H., & Buus, S. R.** (1988). Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. *The Journal of the Acoustical Society of America, 84*, 195–203.
- Fillgrabe, C., Moore, B. C. J., & Stone, M. A.** (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Frontiers in Aging Neuroscience, 6*, 347.
- Gatehouse, S., & Noble, W.** (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology, 43*, 85–99.
- George, E. L. J., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., & Houtgast, T.** (2007). Auditory and nonauditory factors affecting speech reception in noise by older listeners. *The Journal of the Acoustical Society of America, 121*, 2362–2375.
- Gifford, R. H., Bacon, S. P., & Williams, E. J.** (2007). An examination of speech recognition in a modulated background and of forward masking in younger and older listeners. *Journal of Speech, Language, and Hearing Research, 50*, 857–864.
- Gordon-Salant, S.** (2001). Sources of age-related recognition difficulty for time-compressed speech. *Journal of Speech, Language, and Hearing Research, 44*, 709–719.
- Grant, K. W., Summers, V., & Leek, M. R.** (1998). Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 104*, 1051–1060.
- Griffin, D., & Lim, J.** (1984). Signal estimation from modified short-time Fourier transform. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 32*, 236–243.
- Grose, J. H., Mamo, S. K., Buss, E., & Hall, J. W.** (2015). Temporal processing deficits in middle age. *American Journal of Audiology, 24*, 91–93.
- Gunstad, J., Paul, R. H., Brickman, A. M., Cohen, R. A., Arns, M., Roe, D., & Gordon, E.** (2006). Patterns of cognitive performance in middle-aged and older adults: A cluster analytic examination. *Journal of Geriatric Psychiatry and Neurology, 19*, 59–64.
- Healy, E. W., & Bacon, S. P.** (2002). Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing. *Journal of Speech, Language, and Hearing Research, 45*, 1262–1275.
- Helfer, K. S.** (2015). Competing speech perception in middle age. *American Journal of Audiology, 24*, 80–83.
- Henry, B. A., Turner, C. W., & Behrens, A.** (2005). Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners. *The Journal of the Acoustical Society of America, 118*, 1111–1121.
- Herrmann, B., Henry, M. J., Johnsrude, I. S., & Obleser, J.** (2016). Altered temporal dynamics of neural adaptation in the aging human auditory cortex. *Neurobiology of Aging, 45*, 10–22.
- Houtgast, T., & Festen, J. M.** (2008). On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *International Journal of Audiology, 47*, 287–295.
- Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., & Chang, E. F.** (2016). Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. *The Journal of Neuroscience, 36*, 2014–2026.
- Humes, L. E., & Dubno, J. R.** (2010). Factors affecting speech understanding in older adults. In S. Gordon-Salant, R. Frisina, A. Popper, & R. Fay (Eds.), *The aging auditory system. Springer handbook of auditory research* (Vol. 34, pp. 211–257). New York, NY: Springer.
- Humes, L. E.** (2002). Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *The Journal of the Acoustical Society of America, 112*, 1112–1132.
- IEEE.** (1969). IEEE recommended practices for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics, 17*, 227–246.

- Kates, J. M., & Arehart, K. H. (2005). Coherence and the speech intelligibility index. *The Journal of the Acoustical Society of America*, *117*, 2224–2237.
- Kochkin, S. (2010). MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing. *The Hearing Journal*, *63*, 19–20.
- Kuzniarz, J. (1973). Hearing loss and speech intelligibility in noise. In *Proceedings of the IC BEN Congresses on Noise as a Public Health Problem* (pp. 57–71). Dubrovnik, Yugoslavia: International Commission on Biological Effects of Noise.
- Lin, F. R., Niparko, J. K., & Ferrucci, L. (2011). Hearing loss prevalence in the United States. *Archives of Internal Medicine*, *171*, 1851–1853.
- Lin, F. R., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T. B., Purchase-Helzner, E., & Simonsick, E. M. (2013). Hearing loss and cognitive decline in older adults. *JAMA Internal Medicine*, *173*, 293–299.
- Mandel, M. I., Yoho, S. E., & Healy, E. W. (2016). Measuring time-frequency importance functions of speech with bubble noise. *The Journal of the Acoustical Society of America*, *140*, 2542.
- McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B (Methodological)*, *42*, 109–142.
- Mehraei, G., Gallun, F., Leek, M. R., & Bernstein, J. G. (2014). Spectrotemporal modulation sensitivity for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *136*, 301–316.
- Moore, B. C. J. (1996). Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear and Hearing*, *17*, 133–161.
- Moore, B. C. J. (2016). Effects of age and hearing loss on the processing of auditory temporal fine structure. *Advances in Experimental Medicine and Biology*, *894*, 1–8.
- Oxenham, A. J., & Bacon, S. P. (2003). Cochlear compression: perceptual measures and implications for normal and impaired hearing. *Ear and Hearing*, *24*, 352–366.
- Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., & Chang, E. F. (2012). Reconstructing speech from human auditory cortex. *PLoS Biology*, *10*, 175.
- Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *The Journal of the Acoustical Society of America*, *63*, 533–549.
- Plomp, R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech and Hearing Research*, *29*, 146–154.
- Plomp, R., & Duquesnoy, A. J. (1982). A model for the speech-reception threshold in noise without and with a hearing aid. *Scandinavian Audiology. Supplementum*, *15*, 95–111.
- Plomp, R., & Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *The Journal of the Acoustical Society of America*, *66*, 1333–1342.
- Presacco, A., Simon, J. Z., & Anderson, S. (2016). Evidence of degraded representation of speech in noise, in the aging mid-brain and cortex. *Journal of Neurophysiology*, *116*, 2346–2355.
- Profant, O., Škoch, A., Balogová, Z., Tintěra, J., Hlinka, J., & Syka, J. (2014). Diffusion tensor imaging and MR morphometry of the central auditory pathway and auditory cortex in aging. *Neuroscience*, *260*, 87–97.
- Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology*, *54*, 35–54.
- Schneider, B. A., Pichora-Fuller, K., & Daneman, M. (2010). Effects of senescent changes in audition and cognition on spoken language comprehension. In S. Gordon-Salant, R. Frisina, A. Popper, & R. Fay (Eds.), *The aging auditory system. Springer handbook of auditory research* (Vol. 34, pp. 167–210). New York, NY: Springer.
- Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, *5*, 340–348.
- Shen, Y. (2014). Gap detection and temporal modulation transfer function as behavioral estimates of auditory temporal acuity using band-limited stimuli in young and older adults. *Journal of Speech, Language, and Hearing Research*, *57*, 2280–2292.
- Singh, N. C., & Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *The Journal of the Acoustical Society of America*, *114*, 3394–3411.
- Snell, K. B. (1997). Age-related changes in temporal gap detection. *The Journal of the Acoustical Society of America*, *101*, 2214–2220.
- Souza, P. E., & Boike, K. T. (2006). Combining temporal-envelope cues across channels: Effects of age and hearing loss. *Journal of Speech, Language, and Hearing Research*, *49*, 138–149.
- Stothart, G., & Kazanina, N. (2016). Auditory perception in the aging brain: The role of inhibition and facilitation in early processing. *Neurobiology of Aging*, *47*, 23–34.
- Strouse, A., Ashmead, D. H., Ohde, R. N., & Grantham, D. W. (1998). Temporal processing in the aging auditory system. *The Journal of the Acoustical Society of America*, *104*, 2385–2399.
- Teo, C. H., Vishwanathan, S. V. N., Smola, A. J., & Le, Q. V. (2010). Bundle methods for regularized risk minimization. *Journal of Machine Learning Research*, *11*, 311–365.
- Ter Keurs, M., Festen, J. M., & Plomp, R. (1992). Effect of spectral envelope smearing on speech reception. I. *The Journal of the Acoustical Society of America*, *91*, 2872–2880.
- Ter Keurs, M., Festen, J. M., & Plomp, R. (1993a). Effect of spectral envelope smearing on speech reception. II. *The Journal of the Acoustical Society of America*, *93*, 1547–1552.
- Ter Keurs, M., Festen, J. M., & Plomp, R. (1993b). Limited resolution of spectral contrast and hearing loss for speech in noise. *The Journal of the Acoustical Society of America*, *94*, 1307–1314.
- Turner, C. W., Chi, S.-L., & Flock, S. (1999). Limiting spectral resolution in speech for listeners with sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research*, *42*, 773–784.
- Turner, C. W., Souza, P. E., & Forget, L. N. (1995). Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *97*, 2568–2576.
- Vaden, K. I., Matthews, L. J., Eckert, M. A., & Dubno, J. R. (2017). Longitudinal changes in audiometric phenotypes of age-related hearing loss. *Journal of the Association for Research in Otolaryngology*, *18*, 371–385.
- 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.
- Walton, J., Orlando, M., & Burkard, R. (1999). Auditory brainstem response forward-masking recovery functions in older humans with normal hearing. *Hearing Research*, *127*, 86–94.
- Wingfield, A. (1996). Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory. *Journal of the American Academy of Audiology*, *7*, 175–182.

Appendix

Ordinal Regression Analysis

The Clmg was estimated by identifying regions of the modulation spectrum for which removal systematically predicted poor performance whereas preservation systematically predicted good performance. Each pixel location in the bubbles filter domain—corresponding to a particular component in the 2D modulation spectrum—was considered a separate predictor in a multiple ordinal regression on participant responses (keywords correct). Bundle methods for convex risk minimization with L2 regularization (Teo, Vishwanthan, Smola, & Le, 2010) were used to solve the regression problem. This technique finds a set of weights, w , that minimize the following function:

$$f(w) = 0.5\lambda \cdot \|w\| + R(w),$$

where λ is the regularization parameter, $\| \cdot \|$ indicates the L2 (Euclidean) norm, and R is the risk function. Let us denote the vector of true responses as y , where y can take the value 0–5 (keywords correct), and the vector of predicted responses as \hat{y} . For the ordinal regression problem, we seek some w such that $\hat{y}_i - \hat{y}_j < 0$ whenever $y_i < y_j$. When this relation is not satisfied, we incur a cost $C(y_i, y_j) = y_j - y_i$. Following Teo et al. (2010), we denote M as the number of response pairs (y_i, y_j) for which $y_i < y_j$. The risk function is then defined as

$$R(w) = \frac{1}{M} \sum_{y_i < y_j} C(y_i(w), y_j(w)) \cdot \max(0, 1 + [\hat{y}_i(w) - \hat{y}_j(w)]).$$

We define the error rate in the ordinal regression as the proportion of M for which $\hat{y}_i - \hat{y}_j > 0$. The crucial outcome from the regression is the set of weights, w , which determine the extent to which each pixel location in the modulation spectrum contributes to intelligibility. A large positive weight indicates a large contribution to intelligibility (i.e., the number of keywords correct tends to increase when such pixel locations are retained in the filtered signal), whereas a small positive weight indicates little contribution to intelligibility, and a negative weight indicates a detraction from intelligibility. We refer to the set of weights collectively as a Clmg. An optimal value of the regularization parameter, λ , which minimized the regression error rate, was determined using 10-fold cross-validation.

The regularization parameter λ is a positive value that determines the magnitude of the regularization penalty. Regularization often produces biased estimates of the regression coefficients but reduces the variance in the estimates, producing a better fit to data not yet “seen” by the model. Thus, λ must be tuned to minimize the regression error rate. This tuning would normally be performed separately for each participant, but because we wanted to combine individual participant Clmgs and compare them across listener groups in a second-level analysis (see Second-Level Analysis: Comparison of Clmgs Across Listener Groups section), it was desirable to maintain a single value of λ across all participants. Therefore, tuning was performed by calculating the mean regression error rate across participants for a given value of λ and choosing an optimal λ value such that this quantity was minimized. Specifically, each participant’s data (2D bubbles filters, responses) were split into training (75%) and test (25%) sets, maintaining the balance of response types (i.e., the distribution of keywords correct) across sets. The analysis was restricted to regions of the 2D bubbles filters spanning from 0 to 20 Hz and 0 to 6 cyc/kHz in the modulation spectrum. The resulting bubbles filters (66 × 86 pixels) were vectorized to generate a set of 5,676 predictor variables. The predictor variables were then centered and scaled. During tuning, ordinal regression was performed as described in the Toward Clmgs: Ordinal Regression Analysis section using only the training data, and the regression error rate for each participant was taken to be the average error rate obtained from a balanced 10-fold cross-validation procedure. This was repeated for λ values ranging from 1 to 1,000 in steps of 0.25 on a base-10 logarithmic scale. The errors were then averaged across participants for each λ value, and the optimal λ was taken to be that value that minimized the across-participant mean regression error rate. The optimal λ was determined to be 100, and this value was maintained across all participants. Further examination revealed $\lambda = 100$ was optimal for each listener group separately.
