DOI: 10.61782/fa.2025.0250

FORUM ACUSTICUM
EURONOISE 2025

# CONTRIBUTION OF AUDITORY DETECTION THRESHOLDS TO SPEECH RECOGNITION IN TEMPORALLY AND SPECTRALLY DEGRADED ENVIRONMENTS FOR OLDER ADULTS

**Daniel Fogerty**[1*] **Judy R. Dubno**[2]
[1] Department of Speech and Hearing Science,
University of Illinois Urbana-Champaign, USA
[2] Department of Otolaryngology-Head and Neck Surgery,
Medical University of South Carolina, USA

## ABSTRACT

The influence of room acoustics and environmental noise can lead to masking and degradation of temporal and spectral properties of speech. These environmental factors also contribute to the well-documented large variability in speech recognition, particularly among listeners with hearing loss. Older adults with normal hearing (ONH) or sloping high-frequency hearing impairment (OHI) completed three speech recognition experiments consisting of 15-16 measures of temporally degraded speech with (1) degraded spectral cues, (2) competing speech-modulated noise, and (3) combined degraded spectral cues in speech-modulated noise. Speech was spectrally shaped according to each listener's pure-tone thresholds. Speech recognition thresholds (SRTs) were determined at 50% percent correct recognition. To capture individual differences in auditory detection, principal components analysis was used to summarize the primary variance in detection thresholds from 0.25 to 8 kHz. This component explained an average of 32% and 52% of the variance in SRTs for ONH and OHI listeners, respectively. Further analysis revealed a primary contribution of detection thresholds below 1 kHz for both groups, with low frequency thresholds also differentiating SRTs under different types of distortion for OHI listeners. Results suggest the importance of low-frequency speech cues for glimpsing speech in temporally modulated backgrounds.

**Keywords:** *speech recognition, temporal envelope, glimpsing, aging, hearing loss.*

## 1. INTRODUCTION

Hearing loss declines with age [1] and results in significant reductions in speech understanding [2], particularly in complex acoustic environments involving noise and reverberation. Factors underlying speech recognition in noise may involve components of attenuation and distortion [3]. The attenuation component is well-established in the literature [4], which is primarily related to the audibility of the speech signal, as determined by detection thresholds. The purpose of this study was to assess individual differences in the recognition of degraded speech for older adults with normal hearing or with hearing loss. The consistency of these results was examined across variable listening environments using three studies of recognition of spectrally and temporally degraded speech.

## 2. METHODS

### 2.1 Participants

A total of 41 older adults were included in this analysis: 20 older adults with normal hearing (ONH; 17F, 3M; mean 67 years, 60-74 years) and 21 older adults with hearing loss (OHI; 13F, 7M; mean 72 years, 60-85 years). All participants completed pure-tone threshold testing at octave audiometric frequencies between 0.25-8 kHz. Detection thresholds for the two listener groups are displayed in Figure 1.

[*] ***Corresponding author***: dfogerty@illinois.edu.

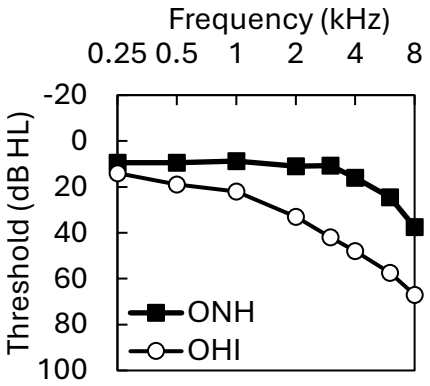


**Figure 1**. Mean audiograms for the two listener groups.

### 2.2 General stimulus processing

All three experiments used temporal-envelope filtered speech selected from the TIMIT or IEEE sentence corpus. Stimuli were bandpass filtered into 18 one-third octave bands. The Hilbert envelopes were extracted from each band. Envelopes were bandpass filtered into two modulation bands: 0–8 Hz and 8–16 Hz. Filtered envelopes were combined with the original spectral components and summed across bands to re-synthesize the original speech sample with reduced temporal modulation cues.

### 2.3 Speech recognition experiments

*2.3.1 Exp. 1: Spectrally reduced speech*

Fifteen acoustic conditions of spectrally reduced speech were analyzed from [5]. These conditions consisted of temporal-envelope filtered speech with additional consonant or vowel intensity scaling to modify the speech modulation depth. The following conditions were tested: (1) two modulation bands (0–8 Hz, 8–16 Hz), (2) two manipulated segments (consonants/vowels), (3) three segment level settings (level factor x0.5, x1.0, x2.0), and (4) three control conditions of the full sentence limited with temporal modulations filtered at 0–8, 8–16, or 0–16 Hz. Sentences were spectrally reduced using a 2 dB signal-to-noise ratio (SNR) signal-correlated noise that preserved temporal modulations.

*2.3.2 Exp. 2: Noise-masked speech*

Sixteen acoustic conditions of temporal-envelope filtered speech and noise were analyzed from Experiment 1 of [6]. These conditions consisted of temporal-envelope filtered speech with additional noise masking using a steady-state noise (SSN) that matched the long-term average spectrum of the target speech, or a speech-modulated noise (SMN) that modulated the SSN by the temporally filtered Hilbert envelope of a different sentence spoken by the target talker. SMN was further processed by expanding or compressing the modulation depth by an exponential factor (K). Four baseline conditions tested included 0-16 Hz temporally filtered speech in unmodulated SSN and in 0-16 Hz SMN at K = 0.5, 1.0, and 2.0. The remaining 12 conditions tested two speech modulation bands (0-8 Hz, 8-16 Hz), in SMN with two noise modulation bands (0-8 Hz, 8-16 Hz), at three noise modulation depths (K = 0.5, 1.0, and 2.0).

*2.3.3 Exp. 3: Spectrally reduced speech + Noise masking*

Sixteen acoustic conditions of temporal-envelope filtered speech and noise were analyzed from Experiment 2 of [6]. This experiment consisted of vocoded speech created during general processing (Sec. 2.2) by combining the filtered Hilbert envelope with the spectral components of the SSN. All other conditions were identical to Exp. 2.

### 2.4 General Procedures

Participants completed all testing in a sound-attenuating booth and listened to stimuli at a sampling rate of 48,828 Hz via one of a pair of Sennheiser HDA 200 headphones following a TDT System III digital-to-analog processor (RP2/RX6) and headphone buffer (HB7/HB5). To ensure audibility of the speech materials (i.e., >15 dB sensation levels) through at least 4.0 kHz, all listeners received frequency-specific gain based on individual detection thresholds (i.e., spectral shaping). Stimuli were presented monaurally to the right ear, unless target sensation levels were closer using the left ear (3 ONH, 14 OHI). To limit the contribution of reduced audibility in the higher frequencies, all stimuli were subsequently passed through a low-pass, linear phase, finite-impulse-response, 128th-order filter with a cutoff of 5.623 kHz. All auditory testing was calibrated to be presented at 70 dB SPL, with a mean presentation level of 82 dB SPL for OHI listeners following spectral shaping.

During speech recognition testing, open-set responses were live-scored and recorded. Participants were encouraged to guess. No feedback was provided. A response was scored as correct if the participant repeated each keyword exactly (e.g., without missing or extra phonemes).

## 3. RESULTS

*3.1.1 Speech recognition thresholds*

Psychometric functions for each of the experiments were obtained by first calculating the degree of speech distortion for each condition using the Extended Short-Time Objective

Intelligibility metric (eSTOI, [7]). The metric compares the spectro-temporal modulation envelopes of the clean and degraded speech signals over short-time segments to produce a similarity measure, with values less than 1.0 indicating the degree of acoustic distortion. From these values logistic functions were fit to the data to determine the 50% point, defining the speech recognition threshold (SRT) for each listener.

*3.1.2 Audiogram factor analysis*

Principal components analysis (PCA) was used to capture the primary variance in audiometric thresholds across both participant groups. Detection thresholds at eight audiometric frequencies (0.25 to 8 kHz) were entered into the analysis to extract factors with eigenvalues greater than 1. A single factor emerged that explained 79.5% of the variance on the full audiogram. All communalities were greater than 0.50 (7 out of 8 were > 0.70, KMO = .89). This analysis resulted in a single audiogram PCA score that captured the primary variance associated with detection thresholds among all listeners.

*3.1.3 Individual differences analysis*

Pearson correlation was used to investigate the relationship between the detection thresholds as predictors of SRTs in the three speech recognition experiments. Figure 2 displays the variance explained by the audiogram PCA score, accounting for an average of 32% (ONH) and 52% (OHI) of the total variance across experiments.

Associations with speech recognition were then examined at octave audiometric frequencies. Results demonstrated higher correlations with speech recognition at the lower frequencies, particularly at 0.25 kHz. Detection thresholds at this frequency were in the normal hearing range (< 20 dB HL, Fig. 1) for both the ONH and OHI groups; these thresholds were not significantly different between groups (p = .11). Contributions were minimal beyond 0.5 kHz for the ONH group. For OHI listeners, detection thresholds accounted for more of the variance for Exp. 2 and 3, which involved listening to speech in speech-modulated noise, than for Exp. 1.

Hierarchical stepwise linear regression was used to predict SRTs for the three experiments using octave detection thresholds followed in a second step by a measure of fluctuating masker benefit (FMB; speech recognition in SMN minus in SSN). Low-frequency thresholds (0.25 kHz) were most predictive for both groups, while higher frequency thresholds (4 kHz) and FMB also contributed for OHI.

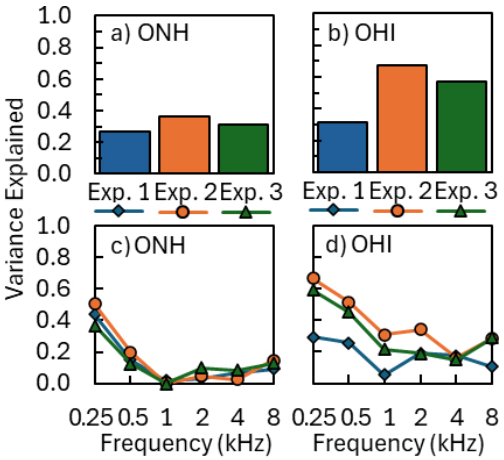


**Figure 2**. Total variance explained ($R^2$) by the (a-b) audiogram PCA and (c-d) detection thresholds at each audiometric frequency for the three experiments (in color) and two listener groups, ONH (left) and OHI (right).

**Table 1.** Hierarchical stepwise linear regression analysis; Additional variance ($R^2$), p < .01

| Group | Predictor | Exp. 1 | Exp. 2 | Exp. 3 |
|---|---|---|---|---|
| ONH | 0.25 kHz | .44 | .50 | .37 |
| OHI | 0.25 kHz | .29 | .66 | .59 |
| | 4 kHz | .18 | .17 | .16 |
| | FMB | .17 | .06 | .06 |
| | *TOTAL* | *.64* | *.89* | *.81* |

## 4. DISCUSSION

Detection thresholds, as summarized by the audiogram PCA, explained a significant proportion of the total variance for both listener groups. This was unexpected because spectral shaping and low-pass filtering were used to ensure adequate speech audibility. Both ONH and OHI groups had some degree of hearing loss at 1 kHz and above, which was captured by the audiogram PCA. Previous work has also identified that hearing loss severity, in this case indexed by the four-frequency pure-tone average, is associated with speech recognition for words and sentences in noise, even after factoring out the contribution of audibility [8]. Thus, detection thresholds appear to capture some component important for temporally/spectrally degraded speech recognition beyond audibility.

Further insight into this relationship is provided by examining detection thresholds at each frequency. Correlations revealed greater contributions of low-frequency hearing, particularly at 0.25 kHz where both groups had normal hearing (thresholds < 20 dB HL; mean speech level

= 42 dB HL). This reflects a general inadequate use of *audible* low-frequency speech cues by ONH and OHI groups, potentially related to suprathreshold differences in processing. Detection thresholds at a broader range of frequencies contributed to speech recognition for the OHI group, including above 1 kHz where they had elevated detection thresholds. Higher associations for OHI were also obtained for Exp. 2 and 3 that involved temporally fluctuating noise. These results suggest the importance of the use of low-frequency speech cues to glimpsing speech, particularly in temporally modulated backgrounds.

The importance of low-frequency cues to recognition of temporally/spectrally degraded speech could potentially be due to the contribution of vocal pitch for speech segregation (e.g., [10]). Other work has highlighted the importance of the use of low-frequency speech cues for speech glimpsing, such as with electro-acoustic hearing (e.g., [11]). This latter study demonstrated an improvement in SNR in the low-frequency band for voiced segments. Thus, a combination of F0 and F1 information from vowels and better glimpsing may contribute to the importance of adequate use of low-frequency cues. The finding of high associations in the present study, even for spectrally reduced speech that may degrade F0 and F1 cues in Experiments 1 and 3, suggests that glimpsing may be the primary contributor to this effect. The higher associations in modulated noise for OHI listeners also supports this view of a glimpse-related mechanism.

Overall, these results highlight the important contribution of low-frequency speech cues to the recognition of degraded speech, and potential suprathreshold differences in using this information by older listeners. Low-frequency cues, such as F0 and F1, may provide critical information to facilitate speech glimpsing in noisy environments.

## 5. ACKNOWLEDGMENTS

Jayne Ahlstrom, Rachel Madorskiy, and Blythe Vickery provided research assistance. This work was supported, in part, by the National Institutes of Health, National Institute on Deafness and Other Communication Disorders, Grants No. R01 DC015465 (D.F.) and R01 DC000184 (J.R.D.), the National Center for Advancing Translational Sciences of the National Institutes of Health under (grant number UL1 TR001450). Some of the research was conducted in a facility constructed with support from Research Facilities Improvement Program (grant number C06 RR 014516) from the National Institutes of Health/National Center for Research Resources.

## 6. REFERENCES

[1] International Organization for Standardization, "Acoustics - Statistical distribution of hearing thresholds as a function of age," ISO 7029:2000, 2000.

[2] U. Hoppe, T. Hocke, and H. Iro, "Age-related decline of speech perception," Front. Aging Neurosci., vol. 14, p. 891202, 2022.

[3] R. Plomp, "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am., vol. 63, no. 2, pp. 533–549, 1978.

[4] L. E. Humes and J. R. Dubno, "Factors affecting speech understanding in older adults," in The Aging Auditory System, S. Gordon-Salant, R. D. Frisina, A. N. Popper, and R. R. Fay, Eds. New York, NY: Springer, 2010, pp. 211–257.

[5] D. Fogerty, J. B. Ahlstrom, and J. R. Dubno, "Sentence recognition with modulation-filtered speech segments for younger and older adults: Effects of hearing impairment and cognition," J. Acoust. Soc. Am., vol. 154, no. 5, pp. 3328–3343, 2023.

[6] D. Fogerty, J. B. Ahlstrom, and J. R. Dubno, "Attenuation and distortion components of age-related hearing loss: Contributions to recognizing temporal-envelope filtered speech in modulated noise," J. Acoust. Soc. Am., vol. 156, no. 1, pp. 93–106, 2024.

[7] J. Jensen and C. H. Taal, "An algorithm for predicting the intelligibility of speech masked by modulated noise maskers," IEEE/ACM Trans. Audio Speech Lang. Process., vol. 24, no. 11, pp. 2009–2022, 2016.

[8] L. E. Humes, "Factors underlying individual differences in speech-recognition threshold (SRT) in noise among older adults," Front. Aging Neurosci., vol. 13, p. 702739, 2021.

[9] L. E. Humes, "Further evaluation and application of the Wisconsin age-related hearing impairment classification system," Am. J. Audiol., vol. 30, no. 2, pp. 359–375, 2021.

[10] J. P. L. Brokx and S. G. Nooteboom, "Intonation and the perceptual separation of simultaneous voices," J. Phonetics, vol. 10, no. 1, pp. 23–36, 1982.

[11] N. Li and P. C. Loizou, "A glimpsing account for the benefit of simulated combined acoustic and electric hearing," J. Acoust. Soc. Am., vol. 123, no. 4, pp. 2287–2294, 2008.