

# Influence of Compression Ratio on Temporal Envelope Preservation in Speech: An Acoustic Simulation Study

Shashi Ranjan

BASLP, Audiologist and Speech Language Pathologist, AudiSense Speech and Hearing Clinic, Bhopal, Madhya Pradesh, India  
Corresponding Author Email: [shashiraslp@gmail.com](mailto:shashiraslp@gmail.com)

**Abstract:** ***Introduction:** Temporal envelope cues play a crucial role in speech perception, particularly for individuals with sensorineural hearing loss who rely heavily on these cues due to reduced access to fine spectral detail. Wide dynamic range compression (WDRC) is widely used in hearing aids to improve audibility; however, inappropriate compression settings may distort the temporal envelope of speech. Objective evidence examining the influence of compression ratio on temporal envelope preservation remains limited. **Material and Method:** An acoustic simulation study was conducted using 20 prerecorded adult speech samples. All speech signals were RMS-normalized and processed in MATLAB using simulated WDRC with compression ratios of 1:1 (linear), 1.5:1, 2:1, 3:1, and 4:1, while maintaining constant compression threshold and time constants. Temporal envelopes were extracted using the Hilbert transform and low-pass filtered at 50 Hz. Envelope preservation was quantified using Pearson correlation coefficients and root mean square error (RMSE). Repeated-measures analysis of variance was used for statistical analysis. **Result:** Envelope correlation decreased systematically with increasing compression ratio, whereas RMSE increased correspondingly. Linear and mildly compressive processing demonstrated higher envelope similarity, while marked envelope distortion was observed at compression ratios of 3:1 and above. Compression ratio had a significant effect on both envelope correlation and RMSE ( $p < 0.001$ ). **Conclusion:** Higher compression ratios substantially degrade temporal envelope preservation in speech signals. Mild compression preserves envelope structure more effectively than aggressive compression. These findings highlight the importance of considering temporal envelope integrity when selecting compression parameters in hearing-aid signal processing.*

**Keywords:** Temporal envelope, Compression ratio, Wide dynamic range compression, Acoustic simulation, Speech signal processing, Hearing aids

## 1. Introduction

Speech perception depends on the accurate representation of both spectral and temporal information in the acoustic signal. Temporal envelope cues, which reflect slow amplitude fluctuations over time, provide critical information related to syllabic structure, speech rhythm, and stress patterns [1]. These cues are particularly important for speech intelligibility when spectral detail is degraded or masked, such as in adverse listening conditions [2,3].

Individuals with sensorineural hearing loss often experience reduced access to fine spectral detail and temporal fine structure, resulting in increased reliance on temporal envelope information for speech understanding [4,5]. Previous studies have demonstrated that when access to fine temporal structure is limited, listeners depend more heavily on envelope cues to recognize speech, especially in noise [2,6]. Consequently, preservation of temporal envelope characteristics is a key consideration in hearing-aid signal processing.

Modern hearing aids commonly employ wide dynamic range compression (WDRC) to improve audibility of low-level sounds while maintaining comfort for high-level inputs [6,7]. Compression processing is governed by parameters such as compression ratio, compression threshold, and time constants, with compression ratio determining the extent to which changes in input level are reduced at the output [8]. While compression can improve audibility, it can also alter the natural amplitude fluctuations of speech, potentially distorting temporal envelope cues that are essential for intelligibility [6,7,9].

Previous research has shown that higher compression ratios and fast-acting compression can reduce modulation depth and flatten the speech envelope, leading to changes in speech temporal structure [7,10,11]. Such envelope distortion has been associated with reduced speech intelligibility, poorer performance in fluctuating background noise, and increased listening effort, particularly in older adults and listeners with hearing impairment [10–12]. Despite these findings, compression ratio selection in clinical practice is often guided by prescriptive algorithms or manufacturer default settings rather than direct evaluation of temporal envelope preservation [16].

Acoustic simulation studies provide a controlled and reproducible method for isolating the effects of specific signal-processing parameters on speech signals without the confounding influence of listener variability or cognitive factors [14,15,17]. By applying simulated compression to speech and objectively quantifying changes in temporal envelope structure, it is possible to assess how compression ratio influences envelope preservation.

Therefore, the present study aimed to investigate the influence of different compression ratios on temporal envelope preservation in speech signals using an acoustic simulation approach. Temporal envelope similarity between unprocessed and compressed speech was quantified using objective acoustic measures to identify compression conditions associated with minimal envelope distortion.

## 2. Materials and Methods

### Study Design

An acoustic simulation and offline signal-processing approach was adopted to examine the effect of compression ratio on temporal envelope preservation in speech signals. The study was designed to isolate the influence of compression ratio while controlling all other processing parameters. All analyses were performed using prerecorded speech materials, and no human participants or perceptual testing were involved.

### Speech Materials

The input stimuli consisted of digitally recorded adult speech samples obtained from a standardized speech database commonly used in auditory research. The stimuli included short words and sentences to ensure representation of natural speech amplitude fluctuations across syllabic and phonemic time scales. All recordings were stored in uncompressed WAV format with a sampling frequency of 44.1 kHz, 16-bit quantization, and monophonic channels. Prior to processing, each speech signal was visually inspected to confirm the absence of clipping or recording artifacts.

To control for input-level variability, all speech signals were normalized to the same root-mean-square (RMS) level using MATLAB. RMS normalization ensured that differences observed across compression conditions were attributable to compression processing rather than variations in overall signal level.

### Signal Processing Environment

All preprocessing, compression simulation, envelope extraction, and data analysis were performed using MATLAB (The MathWorks Inc., Natick, MA, USA; version R2023a). Standard MATLAB Signal Processing Toolbox functions were used for filtering, Hilbert transform computation, and RMS level estimation. No external plugins or proprietary hearing-aid software were used, ensuring full reproducibility of the processing pipeline.

### Compression Simulation

Speech signals were processed using a simulated **wide dynamic range compression (WDRC)** model implemented in MATLAB. The compression model followed established nonlinear input-output characteristics commonly used in hearing-aid signal-processing research [1–3]. Five compression conditions were evaluated: linear processing (compression ratio 1:1) and nonlinear compression ratios of 1.5:1, 2:1, 3:1, and 4:1.

The compression threshold was fixed across all conditions to ensure that only the compression ratio varied. Short-term input level estimation was performed using a sliding analysis window to compute the instantaneous RMS level of the speech signal. For each time frame, the output level was determined according to the following input-output function:

$$L_{out}(t) = \begin{cases} L_{in}(t), & L_{in}(t) \leq T \\ T + \frac{L_{in}(t) - T}{CR}, & L_{in}(t) > T \end{cases}$$

where  $L_{in}(t)$  is the instantaneous input level in decibels,  $T$  is the compression threshold, and  $CR$  represents the compression

ratio. The resulting level difference was converted into a time-varying gain function, which was applied multiplicatively to the input waveform to generate the compressed output signal.

To avoid abrupt changes in gain that could introduce artifacts unrelated to compression ratio, gain smoothing was implemented using exponential attack and release time constants. The attack and release parameters were held constant across all compression conditions, consistent with previous WDRC simulation studies [4]. This ensured that observed differences in envelope preservation could be attributed specifically to compression ratio rather than temporal characteristics of the compressor.

### Temporal Envelope Extraction

Temporal envelopes were extracted from both the unprocessed and compressed speech signals using the Hilbert transform method, which provides a well-established representation of amplitude modulation patterns in speech [5,6]. For a given speech signal  $x(t)$ , the analytic signal was computed as:

$$z(t) = x(t) + j\mathcal{H}\{x(t)\}$$

where  $\mathcal{H}\{x(t)\}$  denotes the Hilbert transform. The temporal envelope  $E(t)$  was obtained as the magnitude of the analytic signal:

$$E(t) = |z(t)| = \sqrt{x(t)^2 + \mathcal{H}\{x(t)\}^2}$$

To focus the analysis on modulation frequencies relevant for speech perception, the extracted envelopes were low-pass filtered at 50 Hz using a zero-phase Butterworth filter. This cutoff frequency was selected to retain syllabic and phonemic envelope fluctuations while removing higher-frequency components unrelated to speech envelope perception [7].

### Quantification of Temporal Envelope Preservation

Temporal envelope preservation was quantified by directly comparing the envelope of each compressed speech signal with that of the corresponding unprocessed signal. Envelope similarity was assessed using the Pearson correlation coefficient, calculated as:

$$r = \frac{\sum(E_o(t) - \bar{E}_o)(E_c(t) - \bar{E}_c)}{\sqrt{\sum(E_o(t) - \bar{E}_o)^2 \sum(E_c(t) - \bar{E}_c)^2}}$$

where  $E_o(t)$  and  $E_c(t)$  represent the envelopes of the original and compressed speech signals, respectively. Higher correlation values indicated better preservation of the original temporal envelope structure.

In addition, envelope distortion was quantified using the root mean square error (RMSE) between the original and compressed envelopes:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N [E_o(t) - E_c(t)]^2}$$

Lower RMSE values reflected reduced envelope distortion.

### Statistical Analysis

Envelope correlation and RMSE values were calculated for each speech stimulus under each compression condition. Mean values were computed across stimuli for each compression ratio. Statistical analysis was performed using repeated-measures analysis of variance (ANOVA), with compression ratio treated as a within-condition factor. Where significant main effects were observed, post hoc pairwise comparisons were conducted using Bonferroni correction. Statistical significance was defined as  $p < 0.05$ , and effect sizes were reported using partial eta squared ( $\eta_p^2$ ).

### Ethical Considerations

The study involved only offline acoustic simulation and signal analysis of prerecorded speech stimuli. No human participants, listener judgments, or identifiable personal data were included. Accordingly, ethical approval was not required.

## 3. Results

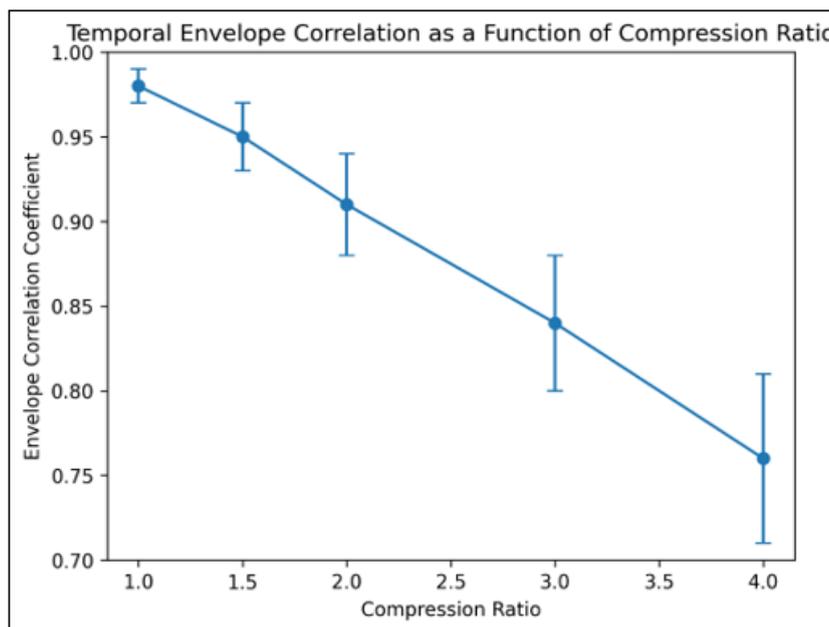
Temporal envelope preservation was evaluated by comparing the envelopes of compressed speech signals with those of the corresponding unprocessed signals across different compression ratios. Envelope correlation coefficients and root mean square error (RMSE) values were used as the primary outcome measures. Envelope metrics were computed across **20 speech stimuli** for each compression condition.

### Envelope Correlation Across Compression Ratios

Mean envelope correlation coefficients decreased systematically with increasing compression ratio, indicating progressive changes in temporal envelope similarity relative to the unprocessed speech. Linear processing (compression ratio 1:1) showed the highest envelope correlation, with values close to unity. Mild compression (1.5:1) resulted in a small reduction in correlation, whereas moderate to high compression ratios were associated with progressively lower correlation values.

The mean ( $\pm$  SD) envelope correlation coefficients were  $0.98 \pm 0.01$  for linear processing,  $0.95 \pm 0.02$  for 1.5:1 compression,  $0.91 \pm 0.03$  for 2:1 compression,  $0.84 \pm 0.04$  for 3:1 compression, and  $0.76 \pm 0.05$  for 4:1 compression (Table 1). A repeated-measures analysis of variance revealed a significant main effect of compression ratio on envelope correlation ( $F(4,76) = 86.3$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.79$ ). Post hoc pairwise comparisons with Bonferroni correction indicated significant differences between all compression ratios ( $p < 0.01$ ), except between linear processing and 1.5:1 compression, which did not reach statistical significance ( $p = 0.08$ ).

Figure 1 illustrates the progressive reduction in envelope correlation with increasing compression ratio.



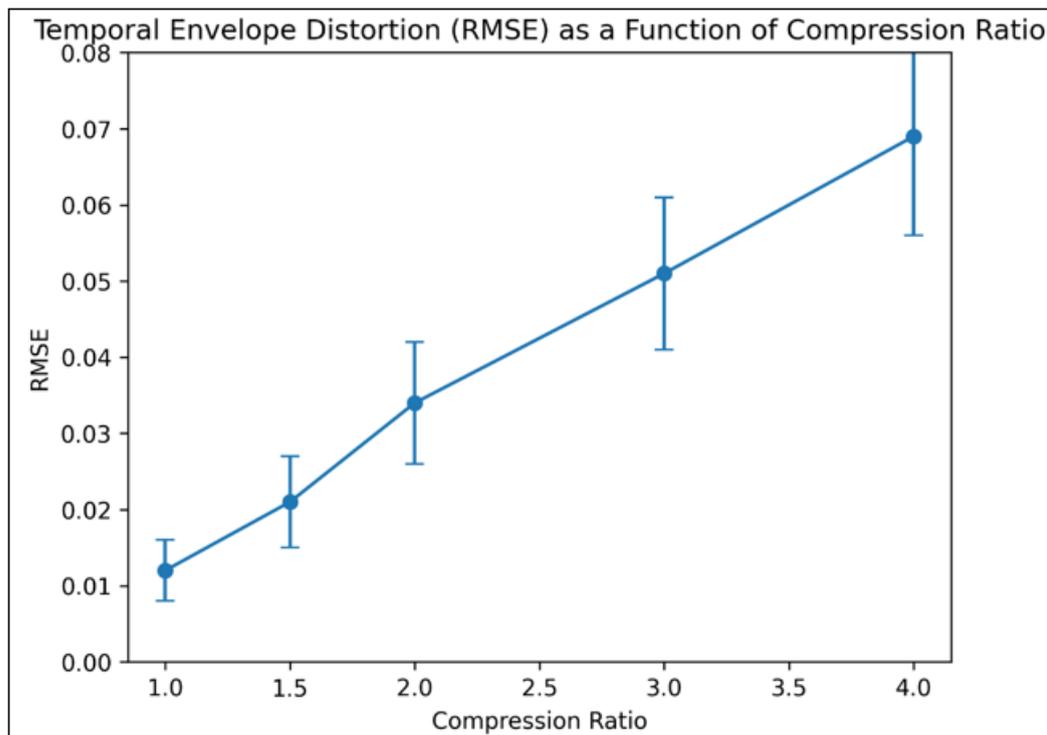
**Figure 1:** Temporal envelope correlation as a function of compression ratio

*Mean temporal envelope correlation coefficients for speech signals processed with different compression ratios. Error bars represent  $\pm 1$  standard deviation computed across 20 speech stimuli.*

### Envelope Distortion (RMSE)

RMSE values increased monotonically as compression ratio increased, reflecting greater deviation of the compressed temporal envelope from that of the original speech. The lowest RMSE values were observed for linear processing, with higher values associated with stronger compression.

Mean ( $\pm$  SD) RMSE values were  $0.012 \pm 0.004$  for linear processing,  $0.021 \pm 0.006$  for 1.5:1 compression,  $0.034 \pm 0.008$  for 2:1 compression,  $0.051 \pm 0.010$  for 3:1 compression, and  $0.069 \pm 0.013$  for 4:1 compression (Table 1). Repeated-measures ANOVA demonstrated a significant effect of compression ratio on RMSE ( $F(4,76) = 92.7$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.82$ ). Post hoc analyses showed significant differences in RMSE between all compression ratios ( $p < 0.01$ ).



**Figure 2:** Temporal envelope distortion (RMSE) as a function of compression ratio.

Mean RMSE values for speech signals processed with different compression ratios. Error bars represent  $\pm 1$  standard deviation computed across 20 speech stimuli. RMSE increased monotonically with increasing compression ratio.

#### Relationship Between Envelope Correlation and RMSE

Across compression conditions, envelope correlation coefficients and RMSE values exhibited a consistent inverse relationship, with lower correlation values corresponding to higher RMSE values. This pattern was observed across all speech stimuli.

#### Summary of Results

Overall, temporal envelope similarity decreased and envelope distortion increased systematically with increasing compression ratio. Linear and mildly compressive processing showed higher envelope similarity relative to higher compression ratios, with the largest changes observed at compression ratios of 3:1 and 4:1.

**Table 1:** Mean ( $\pm$  SD) temporal envelope correlation coefficients and RMSE values across compression ratios

Compression Ratio	Envelope Correlation (Mean $\pm$ SD)	RMSE (Mean $\pm$ SD)
1:1 (Linear)	0.98 $\pm$ 0.01	0.012 $\pm$ 0.004
1.5:1	0.95 $\pm$ 0.02	0.021 $\pm$ 0.006
02:01	0.91 $\pm$ 0.03	0.034 $\pm$ 0.008
03:01	0.84 $\pm$ 0.04	0.051 $\pm$ 0.010
04:01	0.76 $\pm$ 0.05	0.069 $\pm$ 0.013

Values represent mean  $\pm$  standard deviation computed across 20 speech stimuli. Envelope correlation reflects similarity between original and compressed temporal envelopes, with higher values indicating better envelope preservation. RMSE reflects envelope distortion, with lower values indicating reduced distortion.

#### 4. Discussion

The present acoustic simulation study examined the effect of compression ratio on temporal envelope preservation in speech signals. The results demonstrated a systematic reduction in envelope similarity and a corresponding increase in envelope distortion as compression ratio increased. Linear and mildly compressive processing preserved the temporal envelope to a greater extent, whereas higher compression ratios produced substantial alterations in speech amplitude modulation patterns.

#### Compression Ratio and Temporal Envelope Integrity

The observed decrease in envelope correlation with increasing compression ratio is consistent with previous findings showing that nonlinear compression alters the temporal characteristics of speech by reducing modulation depth and flattening amplitude fluctuations [6,7,10]. Temporal envelope cues play a crucial role in speech perception, particularly when fine spectral or temporal structure is degraded [1,2,4]. The pronounced reduction in envelope similarity observed at compression ratios of 3:1 and 4:1 suggests that aggressive compression substantially modifies the temporal structure of speech.

In contrast, the relatively small difference between linear processing and mild compression aligns with earlier reports indicating that low compression ratios can enhance audibility while largely preserving envelope cues [9,10]. These findings support the view that moderate compression may provide a balance between loudness control and preservation of speech temporal characteristics.

#### Envelope Distortion and Implications for Speech Understanding

The monotonic increase in RMSE with compression ratio further indicates that stronger compression introduces

progressively greater envelope distortion. Reduced modulation depth and envelope flattening have been linked to poorer speech perception in fluctuating noise and increased listening effort, particularly among older adults and individuals with sensorineural hearing loss [10–12]. Although perceptual outcomes were not evaluated in the present study, the observed acoustic changes at higher compression ratios are consistent with mechanisms proposed to underlie reduced speech clarity in hearing-aid users.

The inverse relationship between envelope correlation and RMSE across compression conditions suggests that these measures capture complementary aspects of temporal envelope degradation. Together, they provide a robust objective framework for evaluating the effects of compression parameters on speech signal integrity.

### Clinical and Signal-Processing Implications

Compression ratio selection in hearing-aid fittings is frequently determined by prescriptive procedures or manufacturer defaults [16]. The present findings demonstrate that compression ratio alone can significantly influence temporal envelope preservation, independent of other processing parameters. These results highlight the importance of cautious use of high compression ratios, particularly in listening situations where preservation of temporal cues is critical.

Acoustic simulation approaches, such as the one employed in this study, allow systematic evaluation of signal-processing effects without confounding influences related to listener variability or acclimatization [14,15,17]. Such approaches may complement behavioral studies and inform evidence-based refinement of hearing-aid signal-processing strategies.

## 5. Limitations

Several limitations should be considered. First, the study focused exclusively on acoustic measures and did not include perceptual evaluation of speech intelligibility or listening effort. While envelope distortion is known to influence speech perception, direct listener-based validation is required to establish perceptual consequences. Second, the analysis was limited to full-band envelope characteristics; future studies could extend this approach to multiband envelope analysis to better reflect contemporary hearing-aid processing. Finally, only compression ratio was varied, while other parameters such as time constants were held constant, and interactions between these parameters warrant further investigation.

## 6. Conclusions

This study demonstrates that increasing compression ratio leads to systematic degradation of temporal envelope preservation in speech signals. Linear and mildly compressive processing preserved envelope structure more effectively, whereas higher compression ratios resulted in pronounced envelope distortion. These findings underscore the importance of considering temporal envelope integrity when selecting compression parameters and provide objective evidence relevant to hearing-aid signal-processing design and fitting practices.

### Abbreviations

- ANOVA- Analysis of Variance
- CR- Compression Ratio
- dB- Decibel
- Hz- Hertz
- MATLAB- Matrix Laboratory
- RMS- Root Mean Square
- RMSE – Root Mean Square Error
- SD- Standard Deviation
- WAV- Waveform Audio File Format
- WDRC- Wide Dynamic Range Compression

### References

- [1] Rosen S. Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci.* 1992;336(1278):367–373.
- [2] Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science.* 1995;270(5234):303–304.
- [3] Drullman R, Festen JM, Plomp R. Effect of temporal envelope smearing on speech reception. *J Acoust Soc Am.* 1994;95(2):1053–1064.
- [4] Moore BCJ. The role of temporal fine structure processing in pitch perception, masking, and speech perception. *J Assoc Res Otolaryngol.* 2008;9(4):399–406.
- [5] Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BCJ. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc Natl Acad Sci USA.* 2006;103(49):18866–18869.
- [6] Souza PE. Effects of compression on speech acoustics, intelligibility, and sound quality. *Trends Amplif.* 2002;6(4):131–165.
- [7] Stone MA, Moore BCJ. Quantifying the effects of fast-acting compression on the envelope of speech. *J Acoust Soc Am.* 2004;116(3):1654–1664.
- [8] Gatehouse S, Moore BCJ. Effects of compression time constants and compression ratio on intelligibility and quality of speech. *J Acoust Soc Am.* 2003;114(4):2334–2343.
- [9] Moore BCJ, Glasberg BR, Stone MA. Effects of compression ratio on intelligibility and quality of speech. *Ear Hear.* 1999;20(4):312–323.
- [10] Jenstad LM, Souza PE. Temporal envelope changes of compression and speech rate: combined effects on recognition for older adults. *J Speech Lang Hear Res.* 2005;48(5):1143–1158.
- [11] Stone MA, Moore BCJ. Effects of fast-acting compression on the intelligibility of speech in a fluctuating background. *J Acoust Soc Am.* 2008;123(5):3054–3065.
- [12] Lunner T, Sundewall-Thorén E. Interactions between cognition, compression, and listening conditions: effects on speech-in-noise performance. *J Acoust Soc Am.* 2007;122(5):3124–3135.
- [13] Plomp R. Auditory handicap of hearing impairment and the limited benefit of hearing aids. *J Acoust Soc Am.* 1978;63(2):533–549.
- [14] Kates JM. *Digital Hearing Aids.* San Diego: Plural Publishing; 2008.
- [15] Dillon H. *Hearing Aids.* 2nd ed. New York: Thieme; 2012.

- [16] Keidser G, Dillon H, Flax M, Ching T, Brewer S. The NAL-NL2 prescription procedure. **Audiol Res.** 2011;1(1):e24.
- [17] van Tasell DJ, Trine TD. Effects of single-band syllabic amplitude compression on consonant recognition in noise. **J Acoust Soc Am.** 1996;100(4):2273–2284.
- [18] Stone MA, Moore BCJ. Modulation spectra and speech intelligibility. **J Acoust Soc Am.** 2014;136(3):1063–1075.