Effects of Entropy in Real-World Noise on Speech Perception

Erik Jorgensen, AuD, Inyong Choi, PhD, and Yu-Hsiang Wu, MD, PhD

Department of Communication Sciences and Disorders, University of Iowa

Hearing Aid and Aging Research Laboratory

INTRODUCTION

- Traditional tests of speech-in-noise perception use artificial background noise, leading to results that do not generalize to real-world scenarios. Speech perception is poorer in real-world noise than artificial noise, but why this is the case is unclear (e.g., Best et al., 2015).
- There is a significant need for improved ecological validity in laboratory and clinic speech-in-noise tests, particularly with respect to the types of noise used. However, the acoustic complexity and random nature of real-world noise poses challenges to its use and the interpretation of results.
- **The purpose of this study was to quantify complexity in real-world noise and measure its effects on speech perception in listeners with normal hearing.**
- For this study, complexity in real-world noise was quantified using entropy. Entropy is a standard measure of complexity (Shannon, 1948), and has known effects on pure tone discrimination (Lutfi, 1993) and subjective perceptions of speech perception in real-world environments (Ghozi et al., 2015). The objective effects of entropy in real-world noise on speech perception have not been systematically evaluated.
- Entropy quantifies the complexity, or amount of information in a signal, as a function of the statistical structure of the signal. The standard formula (Shannon, 1948) for entropy is:

$$H(x) = -\sum_{i=1}^{n} p(xi) \log_2 p(xi)$$

where *p(xi)* is the probability of the *i*th event in signal *x*.

- **We hypothesized that higher entropy variance in the time and frequency domains would yield systematically better speech perception.** Because high entropy variance correlates with low mean entropy (see Figure 1), we also hypothesized that lower mean entropy would result in better speech perception.

METHOD

- Real-world noise stimuli came from the Ambisonics Recordings of Typical Environments Database (Weisser et al., 2019). Eight environments were used: Café 1, Café 2, Church 2, Dinner Party, Food Court 1, Food Court 2, Street Balcony, and Train Station. Target speech sentences were IEEE sentences.
- Original ARTE recordings were decoded to 8-channels, played through an 8-speaker array, and recorded from a KEMAR. Entropy in the time (energy entropy) and frequency (spectral entropy) domains was then systematically analyzed from the binaural recordings.
- Energy entropy was analyzed using a similar method as in Pikrakis et al. (2008) and Giannakopoulos & Pikrakis (2014), and spectral entropy was calculated using the same method as in Misra et al. (2004). For both energy and spectral entropy, calculations were made using Hamming windows with a length of 0.03s and a step size of 0.01s.
- The short-time entropy in the time and frequency domains was calculated across all potential stimuli segments extracted from the ARTE Database. Standard deviations of entropy were used as a mid-term statistic (e.g., Ghozi et al., 2015; Pikrakis et al., 2008) to quantify the entropy across 3.44s segments, the longest length of a potential target sentence. Examples of the energy (top) and spectral (bottom) entropy sequences for a noise segment (Café 1 at the 125th second), with white noise and a pure tone given for reference, are shown in Figure 1.
- Twenty-five noise segments representing the distribution of energy and spectral entropy variance within environment were identified as the noise stimuli of interest (Figure 2).
- Speech perception in each noise segment was tested in a trial-by-trial design with 400 total trials across two blocks.

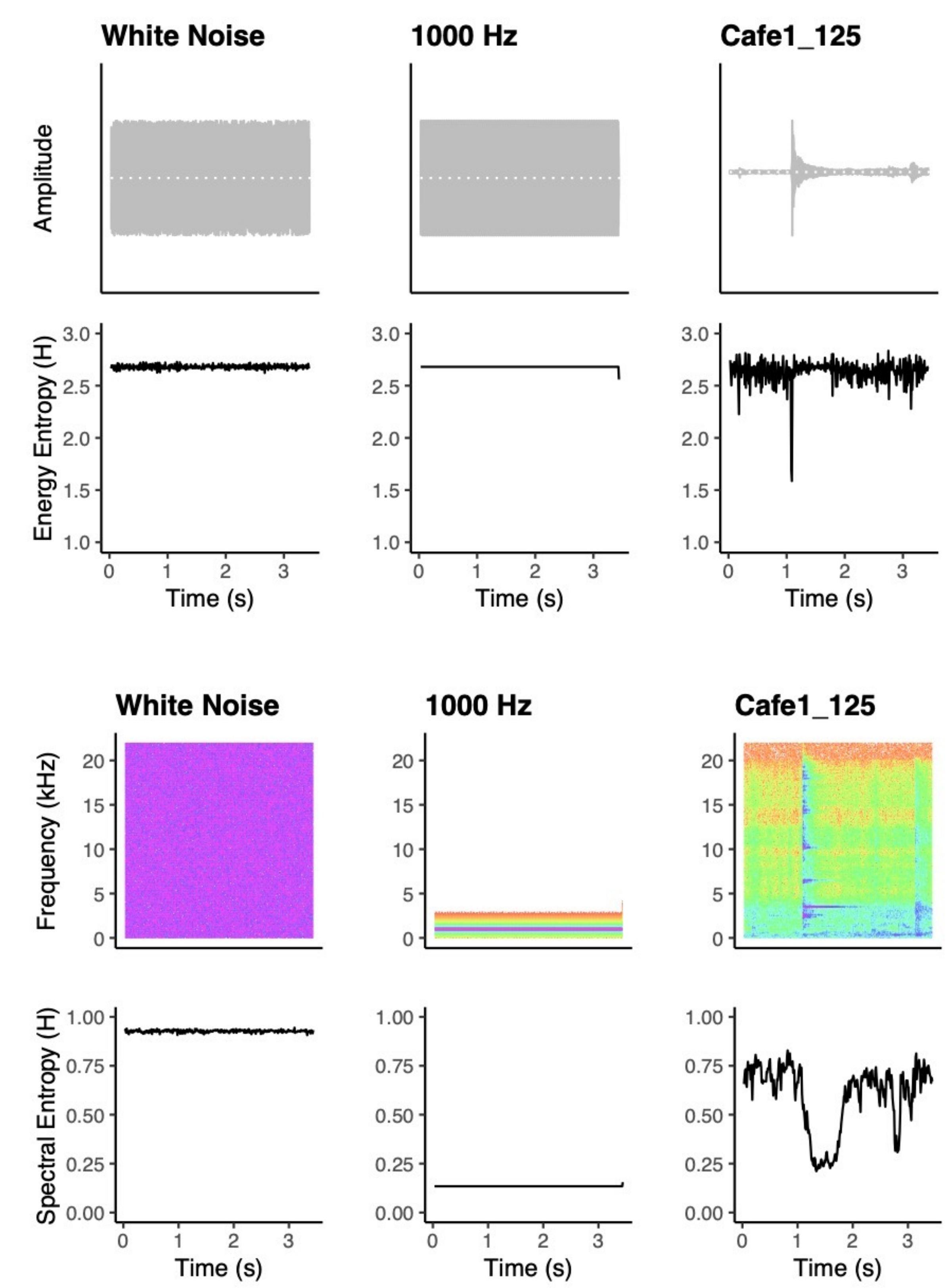

Figure 1. Examples of entropy sequences in the time domain (top) and frequency domain (bottom) for white noise, a 1000 Hz pure tone, and a stimuli noise segment.

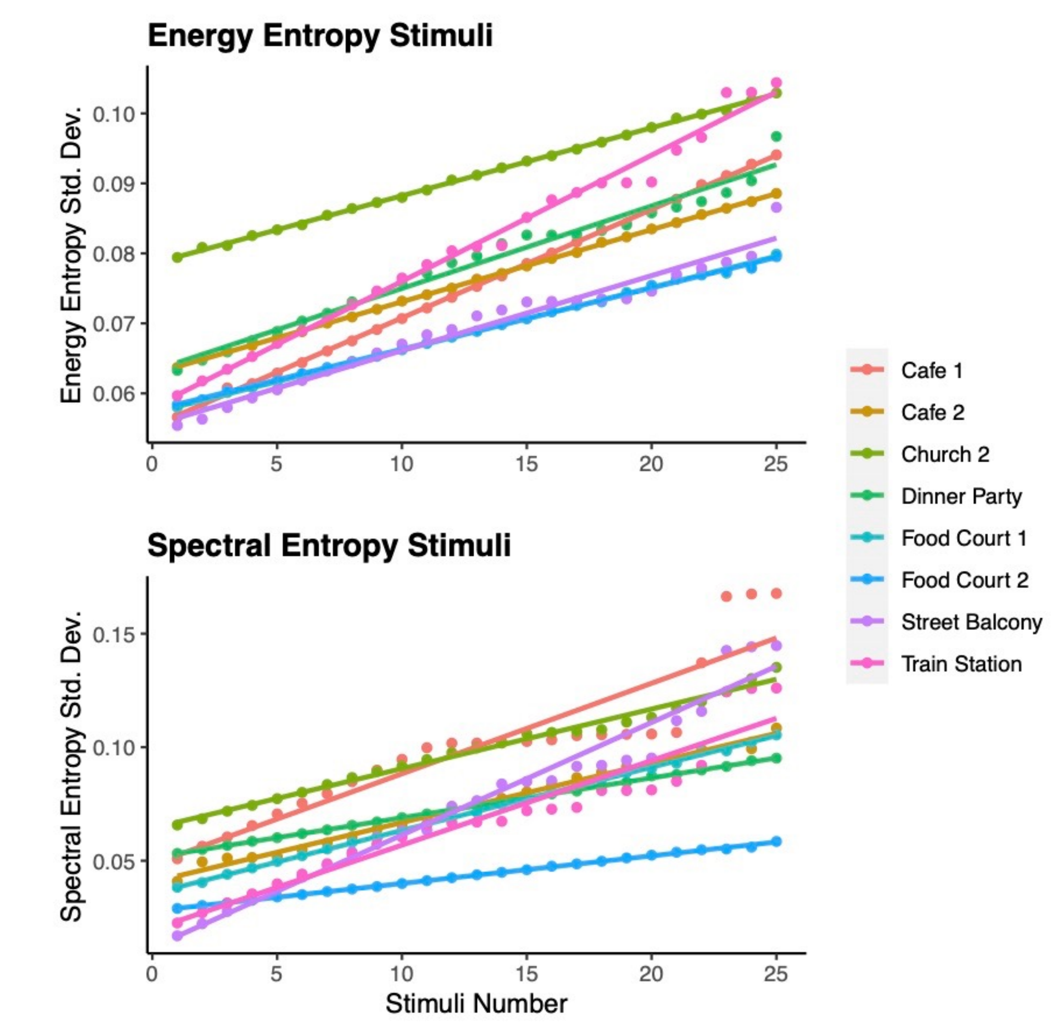

Figure 2. Entropy variance in the time domain (top) and frequency domain (bottom) for all noise stimuli.

- For each participant, target IEEE sentences were drawn randomly, matched with a noise segment in random order, convolved with the room impulse response for that noise segment, and then combined with the noise at -6 dB signal-to-noise ratio. Noise was presented at its real-world level. Noise began 2s before the sentence.
- Scores for each trial were number of keywords repeated back correctly.
- Participants were nine adults with normal hearing (audiometric thresholds <25 dB HL from .25-8 kHz). Ages ranged from 19-54 years (mean = 31 years). All were native English speakers.

RESULTS

- Results showing number of keywords correct as a function of entropy variance in the time (left) and frequency (right) domains are shown in Figure 3. **Number of words correct improved systematically with increases in entropy variance in the time and frequency domains.** Because the baseline entropy of noisy environments tends to be high and variance typically manifests as drops in entropy, the means and standard deviations for both energy (r=.85) and spectral (r=.64) entropy were strongly correlated. A different way to look at the results is that higher mean entropy yielded lower speech scores.
- Linear mixed effects models with random intercepts for participants were used to model the effects of entropy variance on number of words correct. **There was a significant effect of energy (β_1=37.48; t(1790)=11.02, p<.0001) and spectral (β_1=14.74; t(1790)=11.14, p<.0001) entropy variance on number of words correct.** Similar results were observed when analyzing results using energy (β_1=-46.97; t(1790)=-11.2, p<.0001) and spectral (β_1=-7.01; t(1790)=-8.08, p<.0001) entropy mean.

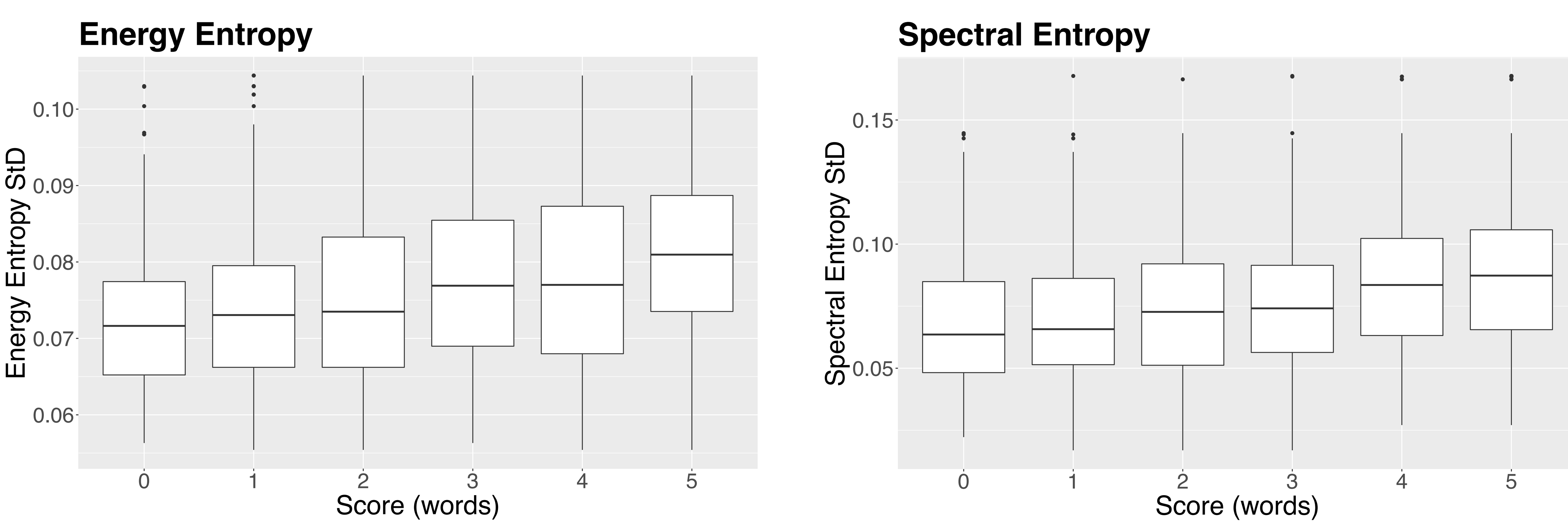

Figure 3. Effect of energy (right) and spectral (left) entropy variance on number of keywords correct across all environments. Horizontal bars represent median values. Vertical bars represent values within the first and third quartiles ± the interquartile range × 1.5. Dots represent outliers.

DISCUSSION

- **Results were consistent with our hypotheses. Increasing noise complexity, quantified with entropy in the time or frequency domains, resulted in systematically poorer speech perception scores.**
- Investigations of speech perception in real-world noise, either in virtual sound environments or real-world environments, should consider the effects of entropy in the design of experiments and interpretation of results.
- This experiment was not reductionist; it is not possible to precisely identify the mechanisms by which entropy affects speech perception in real-world noise. Based on prior work, possibilities include reductions in informational masking with decreasing entropy (e.g., Lutfi 1993), temporal masking release with increases in energy entropy variance (e.g., Miller, 1947), or larger divergences between probability structures of target and masker with increasing entropy variance (Lufti et al., 2013). Central processing and executive function mechanisms may also contribute.
- Entropy variance in real-world noise may account for differences observed between speech perception in laboratory noise and real-world noise, as well as differences in benefit observed from hearing aids in the lab and the real-world (Best et al., 2015; Wu et al., 2019).
- An important future direction for this line of research is an investigation of the effects of entropy in real-world noise on speech perception in listeners with hearing loss, who typically show less benefit from masking release mechanisms than listeners with normal hearing (e.g, Best et al., 2011). An additional critical area of future work is a characterization of how amplification, compression, and other signal processing features, particularly adaptive features, interact with entropy in real-world noise to affect speech perception and hearing aid benefit for listeners with hearing loss.

References: Best et al. IJA. 2015. 54(10), 682–690. Best et al. JASA. 2011. 129(3), 1616–1625. Ghozi et al. JAES. 2015. 63, 475–487. Giannakopoulos & Pikrakis, A. 2014. Lutfi, R. A. JASA. 1993. 94, 748–758. Lutfi, et al. JASA. 2013. 134(3), 2160–2170. Misra et al. IEEE. 2004. 1, 1–193. Pikrakis et al. 33rd International Conference on Acoustics, Speech, and Signal Processing, 2008. 24. Shannon, C. E. The Bell System Technical Journal, 1984. 27(3), 379–423. Weisser, et al. AAUA, 2019. 105(4), 695–713. Wu et al. E&H. 2019. 40(4). **Funding:** NIH/NIDCD F32DC018980 (Jorgensen). **Contact:** Erik Jorgensen, University of Iowa, erik-j-jorgensen@uiowa.edu.