The role of speech sonority in speech perception in a multi-talker environment

This study concerns listeners' ability to understand natural language speech in difficult listening conditions. Previous studies by McAuley et al [4;5] have examined the effect of rhythm on listeners' selective attention in a multi-talker environment. Participants listened to spoken sentences of the form "Ready [Call sign] go to [Color] [Number] now" and reported the Color and Number spoken by a target talker. This frame is called the coordinate response measure (CRM) in the literature, in which four color words are involved [1;2]. There were four levels of rhythm alteration: m=0 (unaltered), m=0.25; m=0.50 (intermediate), m=0.75 (maxim). It was found that listeners' performance became worse when altering the rhythm of the target talker, while keeping the rhythm of the background intact (i.e., "target rhythm effect"). By contrast, altering the rhythm of the background talkers, while keeping the target rhythm intact, improved listeners' performance (i.e., "background rhythm effect"). The findings provide evidence for the selective entrainment hypothesis, which argues that entrainment to the natural rhythm of target speech is proposed to help track target over time while ignoring noisy background.

The current study examines what role sonority plays in both interfering and selectively attending to the target words. Hearers benefit from the perceptual prominence cues of highly sonorous target sounds, but in the meantime they can also get interfered by highly sonorous masker sounds. Sonority influences lexical access through the sonority sequencing principle (SSP), which facilitates the recognition and segmentation of syllables within speech [3;6]. The more non-sonorous the onset of a syllable is, the larger is the degree of sonority rise to the nucleus, the more optimal the SSP [3]. Hence, built up on McAuley et al and SSP, the current study asked the questions of how sonority modulates both the target and the background rhythm effect, namely, to what extent the target and the background rhythm effect could still be observed if we adjust the sonority scale of the target word.

We investigated the role of sonority in listeners' selective attention to a target talker presented in a masker talker male background, with the corrected randomization across subjects. Seven Native speakers of American English, tested normal hearing (PTA at or below 20 dB), were directed to listen to the target and report the color and number. The color words were classified into two categories based on their sonority developed from intensity. The individual sonority index of each segment is extracted from Parker (2002:235-242) [6], and each color word's sonority index is further derived by the formulation that the rise of a sequence XY is defined as "sonority(Y) – sonority(X)" [3;6]. "blue" and "red" were categorized as *non-sonorous*, given their sonority index ("blue"=11.54; "red"=9.08); while "white" and "green" were *sonorous*, given their sonority index ("white"=25.71; "green"=19.45). Target and masker confusion matrices were generated to measure confusability. Diagonal values of target matrices were extracted as proportion correct (Table 1), and diagonals of masker matrices as proportion intrusion (Table 2).

Our findings suggest that in the CRM paradigm, there is a positive correlation between the background color words intrusion and their sonority (R=0.98; p=0.023). Sonorous color words' interference was strong. The increase of sonority improves participants' performance (R=0.96; p=0.039). They excelled when target words were sonorous. If listeners were to internalize the sonority factor (i.e., scale the sonority and then access their sonority dictionary), they can make a judgment prediction on the word. Moreover, individual confusion matrices were calculated for each level of rhythm manipulation, which were then further separated by four different colors (Figure 1). Altering the background rhythm improved listeners' ability to selectively attend to sonorous color words but not necessarily to non-sonorous ones, while altering the target rhythm reduced their accuracy if the target words were sonorous (Figure 1). Put otherwise, both the target and the background rhythm effects could still be observed, if and only if the sonority scale of the target word is high. The established rhythm effect was confined to *sonorous* word identification, indicating that listeners' ability to use speech rhythm in a competition environment is sensitive to speech sonority.

Table 1 Target confusion matrix

	Participants' response			
Target stimulus	blue	red	white	green
blue	0.34	0.13	0.19	0.34
red	0.13	0.26	0.39	0.22
white	0.08	0.07	0.76	0.09
green	0.08	0.07	0.12	0.72

Table 2 Masker confusion matrix

	Participants' response			
Masker stimulus	blue	red	white	green
blue	0.25	0.15	0.32	0.29
red	0.16	0.21	0.31	0.32
white	0.14	0.05	0.52	0.29
green	0.09	0.13	0.31	0.47

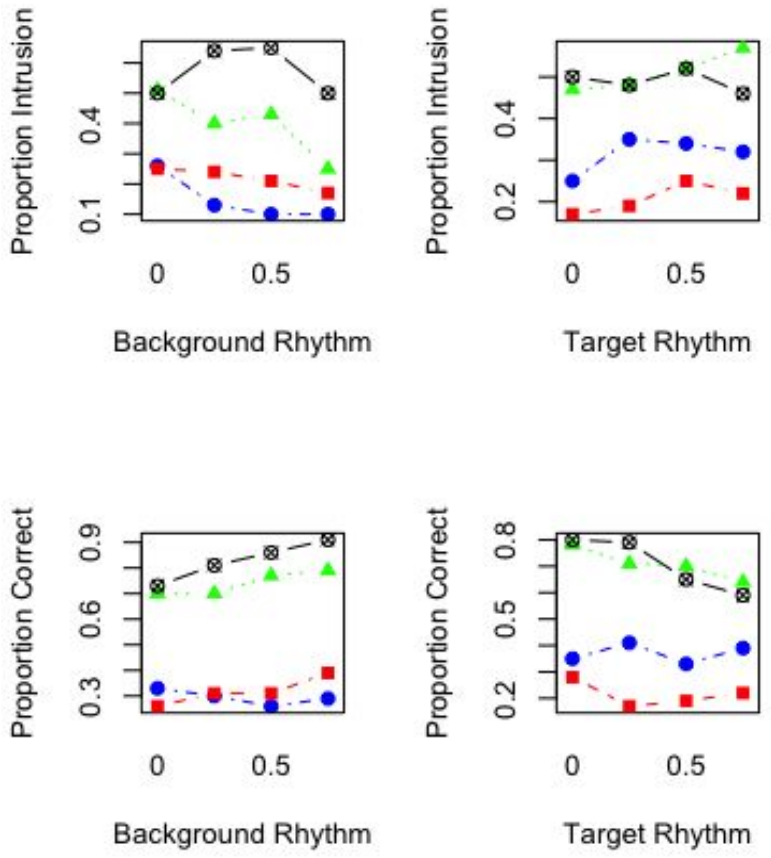

Figure 1 The relationship of rhythm and color words' sonority in CRM paradigm.
x-axis: alteration; lines: color words. Row 1: interference effect; Row 2: hearers' performance

References **[1]** Bolia, R. S. et al. (2000). A speech corpus for multitalker communications research. *Journal of the ASA*, 107(2), 1065-1066. **[2]** Cummins, F. (2015). Rhythm and speech. *The handbook of speech production*,158-177. **[3]** Gordon, M. et al. (2012). Sonority and central vowels: A cross-linguistic phonetic study, In S.Parker (Ed.) *The Sonority Controversy*, 219-256. **[4]** McAuley, J.D. (2010). Tempo and rhythm. *Music Perception*, 165-199. **[5]** McAuley, J.D. et al. (2020). Altering the rhythm of target and background talkers differentially affects speech understanding. *Attention, Perception and Psychophysics* 82, 3222–3233. **[6]** Parker, S. (2002). *Quantifying the sonority hierarchy* Doctoral Dissertation Proquest. AAI3056268.