Confusability and Color Words Perception in Noise Environment

Introduction This study is built up on McAuley et al's (2020) experiment series that investigated listeners' ability to use speech rhythm to attend selectively to a single target talker presented in a single masker talker background. 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. Our results suggest that there is a correlation between the color words that participants tend to choose and the target color word's sonority feature.

Prediction Adopting Hayes (2004, 2009), we classified the color words "blue" "red" "green" and "white" into two groups based their onset consonants sonority: (i) the sonority High group, represented by [+sonorant]; and (ii) the sonority Low group, namely [-sonorant]. The **[+sonorant]** group consists of "white" [wait], given that it has four segments (i.e. four weight) and it has a diphthong /ai/; and "green" [gɹi:n], since it has four segments (i.e. four weight) and /gɹ/ is a complex onset. The **[-sonorant]** group consists of "red" [ɹed], given that it's less heavy (i.e. three weight) and it's a simple syllable; and "blue" [blu:], which has three segments (i.e. three weight) and a long vowel.

From the perspective of vowels quality, we selected the ranking in Hillenbrand (1995), who reports the identification accuracy difference for English vowels: [i] 99%; [u] 97%; [e] 95%. The [i] segment is intrinsically easier to identify, among all the English vowels. In contrast, the [u] segment has low F1 and F2 frequency, which makes it easy to get undermined or even totally covered by the background noise. It's therefore predicted that: (i) in terms of masker, [+sonorant] color words are associated with more background intrusion whereas [-sonorant] means less intrusion; (ii) regarding target, when the color words are [+sonorant], they are more likely to get identified and participants' performance is better. Hence, we calculated both masker and target confusion matrices and plotted them in a similarity space, in order to justify the relation of confusability and color words perception.

Motivation It's worth mentioning that the measure of sonority in this study is based on Hayes (2004), Clements (1988) and Daland et al (2011). Only uncontroversial distinctions are made in this scale, and it represents the consensus of the phonological community: [SCALE] obstruents [/b/, /g/] (0) << nasals << liquids [/l/; /ɹ/] (2) << glides [/w/] (3) << vowels (4).

Various sonority scales have been proposed and discussed in previous studies (Steriade 1982; Selkirk 1984; Clements 1992; Parker 2002; a.o.). They share the properties that: (a) each segment has a sonority value represented by an integer; (b) segments are grouped into sonority classes sharing the same sonority value; (c) the minimally sonorant class has a sonority value of 0; (d) Sonority increments by 1 between different classes (Daland et al 2011:202). To justify the assignment of integer values to different segment classes, we adopted a well-established type of formulation: [FORMULATION] the rise of a sequence XY is defined as sonority(Y) - sonority(X). For example, the onset sequence [bl] is acceptable since sonority(l)-sonority(b) >= 2, and 2 is a commonly seen threshold for acceptable onsets rises. The integer values we assigned to the color words onsets are listed as follows:

[bl] in “blue”: sonority(l)-sonority(b) = 2

[ɹ] in “red”: sonority(ɹ) = 2

[w] in “white”: sonority(w) = 3

[gɹ] in “green”: sonority(ɹ)-sonority(g) = 2

Hayes (2009) independently measured vowels sonority and proposed that low (open) vowels are more sonorant than high (close) vowels, since low vowels are longer and more intense than high vowels. Gordon et al (2012) points out that front vowels (e.g. [i]) have greater auditory energy than back vowels (e.g. [u]) due to the location of the second formant. Therefore, we propose following rank of vowel qualities: [SCALE] [u:] << [e] = [i:] << [a]. To sum up, the consonant (onset) and the vowel qualities are important motivations to group "white" [wait] and "green" [gɹi:n] as [+sonorant], while "blue" [blu:] and "red" [ɹed] as [-sonorant].

Intrusion analysis (1) is an aggregated masker confusion matrix, where the across represents participants' actual response and the vertical refers to the masker. For example, the value 0.16 in $cell_5$ is interpreted as follows: 16% of the time participants reported that they heard "blue", given the masker is "red". The diagonal cells (*1, 6, 11, 16*) show that the word "white" has a salient intrusion ($cell_{11}$ 0.52), whereas "red" and "blue" have a less salient intrusion ($cell_6$ 0.21; $cell_1$ 0.25). This appears to be in line with our claim that "white", which belongs to the sonority High group, is inclined to get a [-sonorant] word undermined. Thus, when the masker is "white", participants tend to respond with "white". But this tendency is less salient when the masker is "blue" or "red", due to the English phonetics fact that they are not very sonorant and their vowels [u] and [e] are not as distinctive as [i] in "white".

Accuracy analysis (2) is an aggregated target confusion matrix, where the across is still the response but the vertical represents the target. For instance, in $cell_9$, 8% of the time when the target is "white", participants respond with "blue". Different from the masker confusion matrix (1), (2) concerns the participants' performance. The diagonal cells are the actual accuracy when the target was the given color, namely the higher the value, the higher the probability that the response and the target match. Thus, we concluded that low sonority words such as "red" are confusable with the rest, while high sonority word “white” is distinct and is less probable to get confused with other segments. That explains why with a “white” target, participants’ performance is better (cell_{11} 0.76).

Similarity space (3a, b) further illustrates the distribution of the color words in a similarity space. (3a) takes in matrix (1) and demonstrates the intrusion from background color words. It suggests that perceptually, “blue” and “red” maskers are clustered, namely they are confusable to each other, whereas “white” and “green” are distinct. By contrast, (3b), which takes in matrix (2), is square-like since it illustrates participants’ attempts to distinguish all the color words and to detect the target. Besides eyeballing the similarity space generated by multidimensional scaling analysis, Pearson’s correlation coefficient analysis was conducted to quantify the relation between confusability and perception. The variables target color words’ sonority and participants’ performance were found to be strongly positively correlated, $r(2) = 1$, $p < .01$ ($t = 19.74$; 95% CI = [0.88,0.99]; sample estimates = 0.997). A scatterplot summarizes the results (4). Increases in rating of sonority were correlated with increases in participants’ responses’ accuracy.

Baseline Importantly, we also consider the baseline scenarios. We calculated the number of trials that occurred for each combination of target/masker, as well as the proportion of times each masker color appeared for a given target color (and vice versa). Everything turns out to be relatively equal. We looked at the counts in terms of percentage and every color is very close to 25%. Specifically, here is an average across all the participants: target color count: “blue” (164) “red” (163) “white” (157) “green” (156); masker number count: “blue” (155) “red” (163) “white” (163) “green” (159). The total count of each color across participants is as follows: blue masker (1085); red masker (1140); white masker (1142); green masker (1113), on the other hand, blue target (1146); red target (1143); white target (1099); green target (1092). (*...add Experiment 4 sanity check?*)

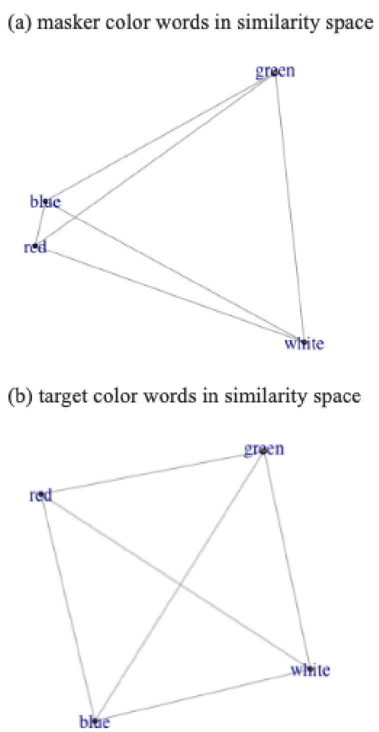
(1) masker confusion matrix

Vertical 👇\Across 👉	blue	red	white	green
blue	[1] 0.25	[2] 0.15	[3] 0.32	[4] 0.29
red	[5] 0.16	[6] 0.21	[7] 0.31	[8] 0.32
white	[9] 0.14	[10] 0.05	[11] 0.52	[12] 0.29
green	[13] 0.09	[14] 0.13	[15] 0.31	[16] 0.47

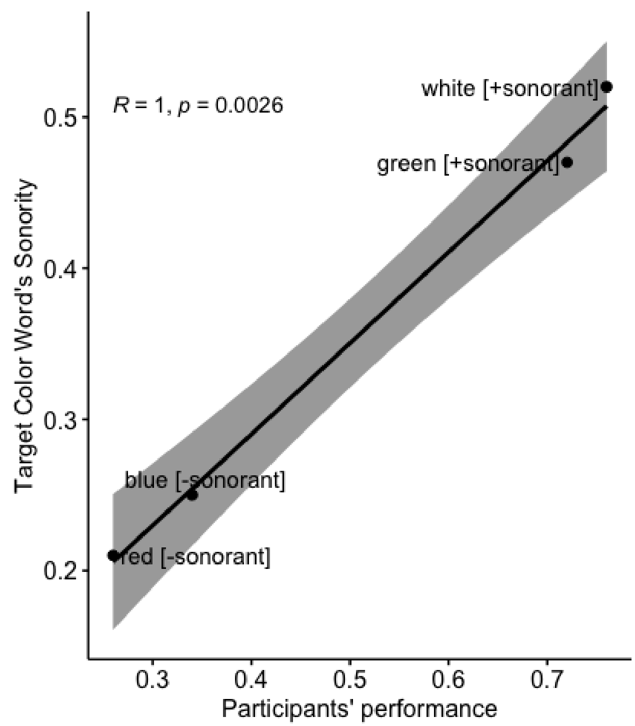
(2) target confusion matrix

Vertical 👇\Across 👉	blue	red	white	green
blue	[1] 0.34	[2] 0.13	[3] 0.19	[4] 0.34
red	[5] 0.13	[6] 0.26	[7] 0.39	[8] 0.22
white	[9] 0.08	[10] 0.07	[11] 0.76	[12] 0.09
green	[13] 0.08	[14] 0.07	[15] 0.12	[16] 0.72

(3) Multidimensional scaling analysis



(4) Pearson correlation - sonority and performance



<u>Discussion</u> This pattern suggests that sonority is a confusable feature, in addition to manner and place of articulation. These results suggest that at least for English, perceptual confusability is argued to be better representations of how speakers judge similarity, rather than patterns in the lexicon or metrics (*such as shared natural classes metric of Frisch et al. (2004)*).

Selected references: Gordon, M. et al. Sonority and central vowels: A cross-linguistic phonetic study, 2012, In *The Sonority Controversy*, edited by Steve Parker, pp. 219-256. Mouton de Gruyter. Hayes, B. (2009). Introductory phonology. Blackwells. Hayes, B et al. (2004). *Phonetically Based Phonology*. Cambridge University Press. McAuley, J.D. et al. Altering the rhythm of target and background talkers differentially affects speech understanding. *Attention, Perception and Psychophysics* 82, 3222–3233 (2020).