Sung Word Recognition is Modulated by Word Frequency and Matching Stress to Musical Prominence

Natsumi Taniguchi (natsumit@usc.edu)[1]

[1]Department of Linguistics, University of Southern California

Background: Sung speech is a naturally occurring type of speech with heavy acoustic perturbations, similar to accented speech or speech-in-noise. Characteristically, the perturbations are caused by pressures from music, a rich mental system that shares cognitive and neural resources with language [5] while also having independent systems and knowledge [2]. Additionally, perception of music without language is known to involve hierarchically organized structures of pitch and rhythm [3]. How musical perturbations in the acoustic signal and perceptions of musical structures independent of language affect the perception of speech has been understudied. One can also assume that linguistic factors influence perception of sung speech similarly to that of spoken speech.

The present study investigates the details of this interaction between music and language in speech perception through a sung word recognition task, manipulating the effect of a linguistic factor that is known to affect lexical access, word frequency, and a musical factor that perturbs the acoustic signal and involves rhythmic perception, stress-prominence matching.

Methods: 61 native English speakers listened to disyllabic words in the carrier sentence, “I am singing the word __ right now,” sung to a simple melody (example in Figure 1), and answered what word they thought was sung. Word frequency of the target word (high vs. low based on log frequency in SUBTLEX [1]) and the alignment between the lexical stress of the target word and metrical prominence of the music were manipulated. Stimuli were sung by 2 trained singers at 150 bpm and were processed with a low-pass filter with a 3k Hz cutoff to avoid ceiling effect through a slight loss of segmental information.

Results: A mixed-effects logistic regression revealed that higher word frequency and matching stress to metrical prominence had positive effects on recognition accuracy (Figure 2). An analysis of the incorrect responses that participants gave further showed that the stress position of disyllabic word responses was influenced by the position of metrical prominence (Figure 3). An example is responding ”dizzy” to ”disease” being sung with metrical prominence on the first syllable.

Discussion: Matching stress and metrical prominence increased accuracy, and metrical prominence influenced the stress patterns of incorrect responses, suggesting that the participants may have utilized their expectations for English text-setting (positions with metrical prominence are likely to contain stressed syllables [4]) in lexical access in addition to acoustic information. The word frequency effect extending to sung word recognition confirms that sung speech perception utilizes lexical access pathways and resources used in spoken speech perception. Overall, the present study provides experimental evidence suggesting that the perception of sung speech engages the resources and perceptual pathways used in both spoken speech perception and musical perception, including the involvement of metrical prominence structure perception and top-down knowledge of how language is often set to music.

References

[1] Marc Brysbaert, Paweł Mandera, and Emmanuel Keuleers. The word frequency effect in word processing: An updated review. 2018.

[2] Xuanyi Chen, Josef Affourtit, Rachel Ryskin, Tamar I Regev, Samuel Norman-Haignere, Olessia Jouravlev, Saima Malik-Moraleda, Hope Kean, Rosemary Varley, and Evelina Fedorenko. The human language system, including its inferior frontal component in "broca's area," does not support music perception. 2023.

[3] Carolyn Drake, Mari Riess Jones, and Clarisse Baruch. The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. 2000.

[4] John Halle and Fred Lerdahl. A generative textsetting model. 1993.

[5] Daniele Schön, Cyrille Magne, and Mireille Besson. The music of speech: Music training facilitates pitch processing in both music and language. 2004.

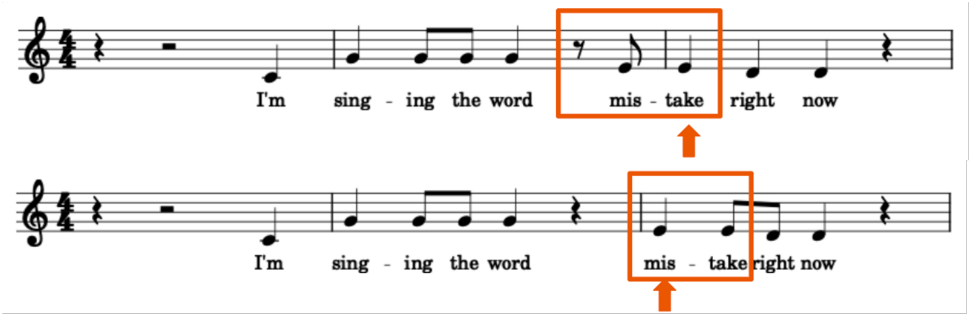

Figure 1: Example stimuli with matching (above) and mismatching (below) lexical stress and metrical prominence. The orange arrow marks the position of metrical prominence. *Mistake* is an example of a final stress, high frequency word.

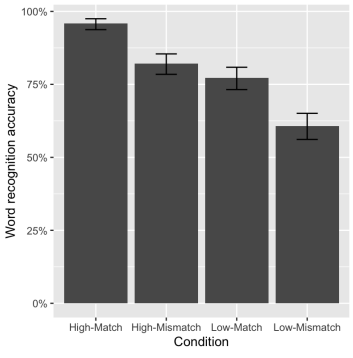

Figure 2: Accuracy of responses between word frequency conditions and stress-prominence matching conditions (1940 trials). Error bars are Clopper-Pearson binomial confidence intervals.

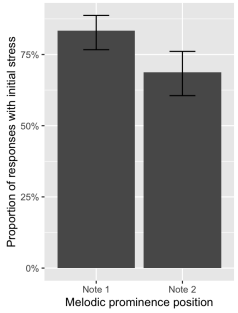

Figure 3: Proportion of disyllabic incorrect responses with initial stress relative to the metrical prominence of the music (309 trials). When melodic prominence is on note 1, there are more initial-stress words in responses. Error bars are Clopper-Pearson binomial confidence intervals.