Title. Across Levels of Representation: Pragmatic Expectations about Prosody Influence the Perception of Loudness

Authors: Moshe Poliak[1], Karenna Caton[1], Anna Zhou[2], Edward Gibson[1]
[1]Massachusetts Institute of Technology [2]Wellesley College

Introduction. To correctly infer the intention of the speaker is a primary goal of language processing. According to Noisy Channel Processing, this is done by integrating the perceived utterance with prior expectations about the language and the world[1-2]. Previously, it has been shown that high-level expectations about plausibility and structure influence the final interpretation of sentences[3-6]. Alternatively, it has been shown that low-level expectations on the word level influence how sounds are perceived, like in the case of phonemic restoration[7-9]. However, it is not clear whether high-level expectations can affect low-level perception. We investigate whether high-level discourse expectations can affect low-level loudness perception, which would suggest that Noisy Channel Processing uses all possible priors to infer the intended input on any level (sounds to meaning), or whether the influence of priors is restricted to their own level of processing. We test this using corrective prosody (e.g., “No, ANYA opened the window”), which is associated with increased loudness of the corrected word (together with increased duration and pitch)[10]. Therefore, if the context informs participants as to which word is being corrected, they should expect that word to be louder than if no such expectation is present. In this study, we manipulated participants’ expectations regarding which word in the sentence is being corrected. We hypothesized that, when a word is expected to be corrected, it is perceived as louder than the same word when there is no such expectation.

Method. *Participants*: We recruited 50 monolingual English speakers from Prolific and excluded 2 for failing to identify the loudest word on more than 80% of the filler items. *Materials:* A native speaker of American English recorded 64 sentences in English (32 critical, 32 filler) of the form “No, {subject} {verb}ed the {object}.” Critical sentences were recorded with neutral prosody while filler sentences had strong corrective prosody on either the subject, verb, or object. Each recording appeared only once per participant in one of four context manipulations: no question, subject, verb, or object (**Table 1**). *Procedure:* participants completed a headphone check[11] before starting the experiment, which consisted of two blocks: one with the *no question* condition and one with the rest (block order was random). After hearing each recording, participants saw the sentence in written form and clicked on the loudest word. In the block with the questions, participants retyped the question before they heard the audio recording.

Results & Discussion. After exclusions, 98.4% of responses correctly identified the loudest word in the fillers. We analyzed the data for critical items (**Figure 1**) using a categorical regression from the nnet[12] package in R, with the response being the word selected as loudest (*subject*, *verb*, *object*), and the predictor being the condition (*no question*, *subject*, *verb*, *object*). We then used the emmeans package[13] to contrast each level of the dependent variable in the no question condition with each level of the dependent variable in the condition that contrasted it, resulting in one comparison: no question vs. contrasted. The comparison showed a small but reliable effect in the predicted direction, such that contrasted words were selected as louder more often than words in the no question condition (β=0.041 (logits), SE=0.013, df=8, t=3.191, p=0.013). We therefore conclude that when a word is expected to be produced with corrective prosody it is perceived as louder than when no such expectation exists; **pragmatic expectations about prosody influence the perception of loudness**. These results suggest

that noisy channel processing uses all types of available priors to infer the intended input across all levels of processing, from sentential meaning to the loudness percept of individual words.

Table 1. Sample questions for each condition for the item “No, Anya opened the window.”

Condition	Question text
no question	-
subject	Did Kelly open the window?
verb	Did Anya close the window?
object	Did Anya open the lid?

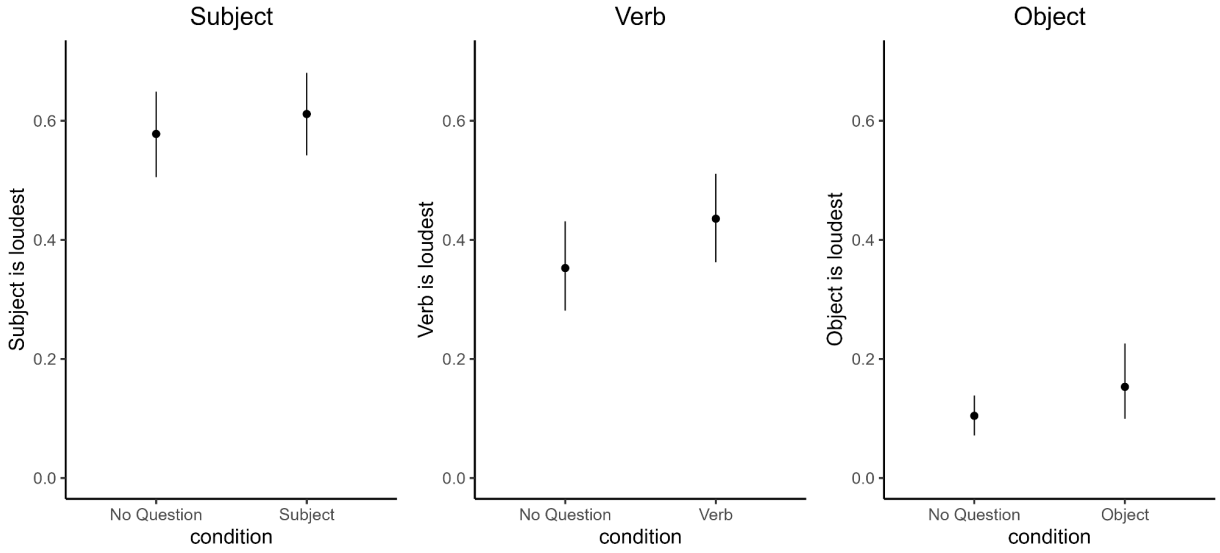

Figure 1: The relative proportion of each word being selected as loudest when it appears in a neutral context (no question) vs. when it is contrasted (see **Table 1**). We computed the proportion of times each part of speech was selected as loudest for each condition and split the data into the three levels of the DV (subject, verb, object). Therefore, the values for “no question” sum up to 1, while the values for the contrasted conditions (subject, verb, object) do not. The dots are overall mean proportions, while the error bars are 95% bootstrapped confidence intervals over item summaries as described above.

References: [1]Shannon (1949); [2]Levy (2008); [3]Gibson et al. (2013); [4]Chen et al. (2023); [5]Poppels & Levy (2016); [6]Poliak et al. (2023); [7]Warren (1970); [8]Cole (1973); [9]Samuel (1981); [10]Breen et al. (2010); [11]Woods et al. (2017); [12]Venables & Ripley (2002); [13]Searle et al. (1980)