Exemplar theoretic modelling of the perception of phrasal prominence

Brooklyn Sheppard[1] and Stephen Winters[1]

[1] University of Calgary

brooklyn.sheppard1@ucalgary.ca

The goal of this study is to investigate the validity of exemplar theoretic computations as a cognitive mechanism for identifying phrasal prominence in English. Exemplar theory, a prominent framework in cognitive psychology, posits that each experience of a given category is stored in the memory of the perceiver, and that when new exemplars of that category are encountered, they are compared to all previous instances of each possible category, with categorization determined by the overall degree of similarity between the new exemplar and all those stored to date. In the realm of linguistics, previous research suggests that exemplar theory can account for a variety of perceptual phenomena at both the segmental and lexical levels of speech [1, 2]. More recently, several studies have begun to test whether this theory can account for perceptual patterns in the prosodic domain as well [3, 4, 5, 6], however, the majority of these efforts have primarily focused on word-level classifications [7, 8]. Prosody, however, doesn't always align with lexical or syllabic units; rather, it often operates at the utterance level, with entire pitch or intensity contours serving as exemplars. Thus, the present study aimed to answer the following research questions: **1)** *Can an exemplar-inspired computational model adequately classify the location of phrasal prominence in statements, yes/no questions, and echo questions using only prosodic characteristics of the whole utterance?* and **2)** *How do the different prosodic cues of F0 and intensity contribute to the classification patterns of an exemplar theoretic model in statements, yes/no questions, and echo questions?* To answer these research questions, ten phonetically trained speakers recorded utterances with prominence on the initial, medial, or final word in statements (e.g., *Liam fed a lamb.*), yes/no questions (e.g., *Did Liam feed a lamb?*), and echo questions (e.g., *Liam fed a lamb?*), which were then used to train and evaluate several exemplar theory-inspired computational models, following [1]. That is, models stored equidistantly sampled points of whole-utterance contours of intensity and/or F0 from the training exemplars in memory for comparison during testing, along with their prominence position label (initial, medial, or final) [9, 10, 11]. Models differed in whether they had access to the intensity contour, the F0 contour, or both for determining a given utterance's location of prominence. To make this determination, each test utterance was compared to all training exemplars by calculating the negative exponent of the Euclidean distance between contours following the original model put forward by [1] with similarity scores summed across all training exemplars of a given category, and subsequently classified as belonging to the category with which it had the highest overall similarity. Results showed that the exemplar models achieved 94.07%, 93.52%, and 90.19% accuracy on categorizing prominence location in declaratives, yes/no questions, and echo questions, respectively. A statistical analysis of accuracies across all computational simulations revealed that declaratives were best classified when only intensity was included as a cue to prominence, while both the yes/no questions and echo questions were best classified when F0 was included as a cue. These results suggest that exemplar models can accurately identify phrase-level prominence using only prosodic features at the utterance-level; they also make predictions for which acoustic cues are most relevant to the perception of phrasal prominence by human listeners. Thus, the present study suggests that exemplar storage and similarity computations may be a powerful framework for explaining the perception of phrasal prominence in English.

References

[1] Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. *Talker variability in speech processing*, 145–165.

[2] Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of experimental psychology: Learning, memory, and cognition*, *22*(5), 1166.

[3] Chow, U. Y. P. (2017). *An exemplar-based model of intonation perception of statements and questions in english, cantonese, and mandarin* [Master's thesis, University of Calgary, School of Languages, Linguistics, Literatures, and Cultures].

[4] Xie, X., Buxó-Lugo, A., & Kurumada, C. (2021). Encoding and decoding of meaning through structured variability in intonational speech prosody. *Cognition*, *211*, 104619.

[5] Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, *46*(2-3), 115–154.

[6] Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in second language acquisition*, *24*(2), 215–221.

[7] Calhoun, S., & Schweitzer, A. (2012). Can intonation contours be lexicalised? implications for discourse meanings. *Prosody and meaning*, 271–328.

[8] Schweitzer, A. (2019). Exemplar-theoretic integration of phonetics and phonology: Detecting prominence categories in phonetic space. *Journal of Phonetics*, *77*, 100915.

[9] Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*(2), 1038–1054.

[10] Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and cognitive processes*, *25*(7-9), 1044–1098.

[11] Bartels, C., & Kingston, J. (1994). Salient pitch cues in the perception of contrastive focus. *The Journal of the Acoustical Society of America*, *95*(5_Supplement), 2973–2973.