The Role of Intonation in Attention Allocation in Serial Recall

Christine T. Röhr[1], *Michelina Savino*[2,1], *T. Mark Ellison*[1], *Martine Grice*[1]
[1]I*f*L Phonetik, University of Cologne, Germany
[2]Department of Education, Psychology, Communication, University of Bari, Italy

In speech processing, prosodic prominence is crucial for directing listeners' attention to the most important parts of the message. For instance, in West-Germanic languages important words typically bear a pitch accent (e.g. [1, 2, 3, 4, 5]). Some recognition memory tasks have shown that prosodically prominent words are recalled faster and more accurately than less prominent words [6]. Moreover, it has been found that recognition accuracy increases when words bear a rising L+H* accent rather than the shallower H* accent (for American English [7] and German [8]). The latter results are compatible with findings attesting that pitch accents with a large f0 rise are perceived as the most prominent (in German [9]), and that intonational rises consume more attentional resources than falls (e.g. [10, 11]).

In languages with both pitch accents and boundary tones, it is generally assumed that pitch accents are the primary encoders of prominence [12, 13]. This suggests that rises on a pitch accent direct listeners' attention more than those at a boundary. However, a study on Italian involving serial recall of nine-digit sequences [14] has shown that rising boundaries at the end of non-final triplets enhance recall, especially on the last items in these triplets. This indicates that boundary tones may also cue prominence.

Here we investigate the effect of rising accents and rising boundary tones on attention in German, using a web-based serial recall task that requires participants to recall sequences of nine digits in the same order in which they are presented. We compared the effect on working memory of sequences grouped by marking the last item of the two non-final triplets with (i) a *rising*-accent + *high*-boundary (accent_RISE), (ii) a *low*-accent + *rising*-boundary (boundary_RISE), or (iii) a *falling*-accent + *low*-boundary (boundary_FALL), as compared to (iv) ungrouped sequences as controls (see Figure 1).

We performed generalised linear mixed-effects models (including prosodic CONDITION as the main fixed factor and assuming random intercepts and slopes for CONDITION by participants) and permutation tests on the recall data of 55 participants (26 female, mean age = 29.2 years, SD = 7.8). Results revealed that all grouped-by-intonation sequences led to better recall performance (i = 78.2 %, ii = 80 %, iii = 77.4 %) than controls (iv = 67.5 %), as a consequence of the grouping effect (likelihood ratio test: significant effect of CONDITION on accuracy, $\chi^2(3) = 55.177$, $p < 0.001$): Figures 2 and 3 show that digits in each position other than the first and last (due to primacy and recency effects) are recalled better in the grouped-by-intonation conditions. Pairwise comparisons confirmed that the three grouped-by-intonation conditions differ significantly from the ungrouped condition (with $p < 0.001$ each), but not from each other. Moreover, a permutation test confirmed that recall accuracy in the ungrouped condition (iv) is significantly ($p < 0.005$) worse for items in all positions except the first two in the sequences, and even in second position (iv) is significantly worse than boundary_RISE (ii) ($p < 0.005$). Results also indicate that intonational rises (i, ii) on non-final triplets led to better recall accuracy than falls (iii). Likelihood ratio tests on the accuracy on the *first and second triplet* registered a significant effect of CONDITION (triplet 1: $\chi^2(3) = 19.989$, $p < 0.001$; triplet 2: $\chi^2(3) = 60.885$, $p < 0.001$) and pairwise comparisons revealed significant differences between the three grouped-by-intonation conditions and the ungrouped condition in both non-final triplets (with $p < 0.001$ each, except for boundary_FALL in triplet 1 with $p < 0.05$), as well as between (ii) the boundary_RISE and (iii) the boundary_FALL conditions in triplet 2

($p < 0.05$). A likelihood ratio test on the accuracy on the *third position in the two non-final triplets* (positions 3 and 6) registered a significant effect of CONDITION ($\chi^2(3) = 50.74$, $p < 0.001$) and pairwise comparisons revealed significant differences between the three grouped-by-intonation conditions and the ungrouped condition (with $p < 0.001$ each), as well as between (i) the accent_RISE and (iii) the boundary_FALL conditions ($p < 0.05$) ((ii) vs. (iii): $p = 0.056$). Although there is no evidence for superior recall on digits bearing accent rises over those with boundary rises, boundary rises appear to facilitate recall over a larger domain, affecting recall of digits within the same triplet: Figure 4 shows improved performance over the whole medial triplet (positions 4, 5, 6) for (ii) the boundary_RISE condition, whereas improvements only occurred on the last item in the same triplet (position 6) for (i) the accent_RISE condition. Permutation tests confirmed a shift up in accuracy from (iii) boundary_FALL to (ii) boundary_RISE at positions 3, 4, 5, 6 ($p < 0.01$), from (iii) to (i) accent_RISE for positions 3 ($p < 0.05$) and 6 ($p < 0.01$), and from (i) accent_RISE to (ii) boundary_RISE at positions 4 ($p < 0.001$) and 5 ($p < 0.05$).

Hence, in general, our results provide evidence that intonational grouping improves recall ability, and that intonational rises lead to better recall performances than falls, indicating that they direct more attention towards an item or even adjacent items. If attention allocation and recall accuracy can be related to prominence, then accentual rises tend to have a local effect on the prominence of the digit, which is accented, whereas boundary rises appear to make the three-digit sequence of the triplet prominent, indicating a more distributed, or global, effect. The global effect of boundary H tones is unsurprising, as in autosegmental-metrical phonology it is analysed as a tone associated with the whole phrase (in this case the three-digit sequence), and could thus be facilitating recall of all items in that phrase, as opposed to an accentual H tone, that may have a more local effect, reflecting its association with a single digit. We have thus shown that intonation can serve to highlight a single item or multiple items within a list, leading to improved recall. We have also shown that the nature of the intonational tones (whether accentual or boundary tones) can determine the position and scope of these improvements.

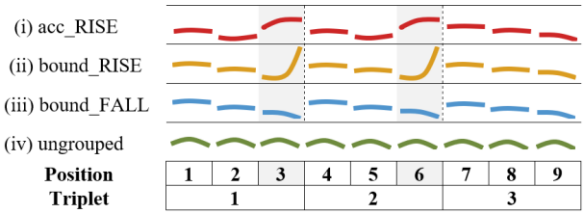

Figure 1. Schematized prosodic patterns of sequence stimuli for the four prosodic conditions. Vertical dashed lines mark intonational group boundaries. Digits in position 3 and 6 (end of first and second triplet) shaded.

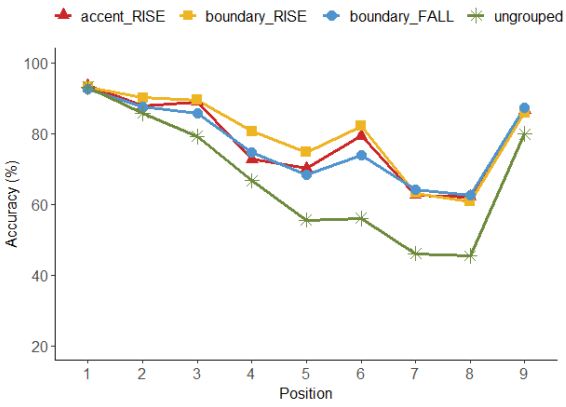

Figure 2. Serial recall curves by condition.

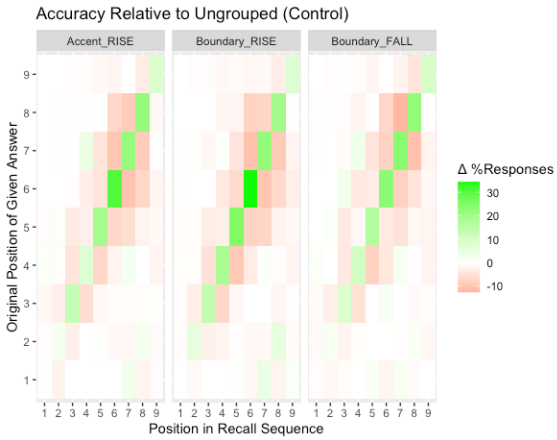

Figure 3. Accuracy of the grouped-by-intonation conditions relative to the ungrouped (control) condition: Position in recall sequence presented (x-axis) against position in recall sequence in responses (y-axis). Off-diagonal values indicate errors.

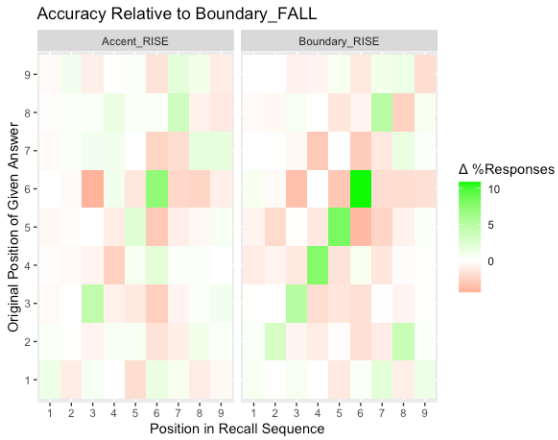

Figure 4. Accuracy of accent_RISE and boundary_RISE conditions relative to the boundary_FALL condition. In the accent_RISE condition improvements are more local (positions 3 and 6), whereas in the boundary_RISE condition improvements are distributed over multiple positions.

References

[1] Pierrehumbert, J. B., & Hirschberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In P. R. Cohen, J. Morgan & M. E. Pollack (Eds.), *Intentions in Communication* (pp. 271-311). Cambridge: MIT Press.

[2] Ito, K., Speer, S., & Beckman, M. (2004). Informational Status and Pitch Accent Distribution in Spontaneous Dialogues in English. *Proceedings of the 2nd International Conference on Speech Prosody* (pp. 279-282). Nara, Japan.

[3] Chen, A., Den Os, E., & De Ruiter, J. P. (2007). Pitch Accent Type Matters for Online Processing of Information Status: Evidence from Natural and Synthetic Speech. *The Linguistic Review, 24*(2-3), 317–344.

[4] Röhr, C. T., & Baumann, S. (2010). Prosodic Marking of Information Status in German. *Proceedings of the 5th International Conference on Speech Prosody* (vol. 100019, pp. 1-4). Chicago, IL, United States.

[5] Grice, M., Ritter, S., Niemann, H., & Roettger, T. (2017). Integrating the Discreteness and Continuity of Intonational Categories. *Journal of Phonetics, 60*, 90-107.

[6] Kember, H., Choi, J., Yu, J., & Cutler, A. (2019.) The Processing of Linguistic Prominence. *Language and Speech, 64*(2), 413-436.

[7] Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition Memory Reveals Just How CONTRASTIVE Contrastive Accent Really Is. *Journal of Memory and Language, 63*(4), 367-386.

[8] Koch, X., & Spalek, K. (2021). Contrastive Intonation Effects on Word Recall for Information-structural Alternatives Across the Sexes. *Memory and Cognition, 49*, 1312-1333.

[9] Baumann, S., & Röhr, C. T. (2015). The Perceptual Prominence of Pitch Accent Types in German. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS XVIII)* (vol. 298, pp. 1–5). Glasgow, UK: The University of Glasgow.

[10] Hsu, C.-H., Evans, J. P., & Lee, C.-Y. (2015). Brain Responses to Spoken F0 Changes: Is H Special? *Journal of Phonetics, 51*, 82–92.

[11] Röhr, C. T., Brilmayer, I., Baumann, S., Grice, M., & Schumacher, P. B. (2020). Signal-Driven and Expectation-Driven Processing of Accent Types. *Language, Cognition & Neuroscience, 36*(1), 33-59.

[12] Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge: Cambridge University Press.

[13] Arvaniti, A. (2020). The Autosegmental-Metrical Theory of Intonational Phonology. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody* (pp. 1-20). Oxford: Oxford University Press.

[14] Savino, M., Winter, B., Bosco, A., & Grice, M. (2020). Intonation Does Aid Serial Recall After All. *Psychonomic Bullettin and Review, 27*, 366-372.