

Intonational cues to item position in lists: evidence from a serial recall task

Michelina Savino¹, Andrea Bosco¹, Martine Grice²

¹Dept. of Education, Psychology, Communication, University of Bari, Italy

²IfL-Phonetics, University of Cologne, Germany

michelina.savino@uniba.it, andrea.bosco@uniba.it, martine.grice@uni-koeln.de

Abstract

Intonation can convey information about how lists are structured into groups, as well as about specific item positions within a group. In Bari Italian, this function is expressed by three different tunes a) a rising contour, signalling that the list has not yet been completed; b) a high-rising contour, marking the penultimate item, i.e. signalling that the end of the list is approaching; c) a falling contour, marking the last item, i.e. cueing the end of the sequence. In this paper we explore the effects of such intonational information on working memory. In particular, we demonstrate that when listeners are requested to recall spoken nine-digit sequences by strictly following their serial order, their performance is significantly better when lists are characterised by tunes of the type described above, compared to sequences whose items are marked by a neutral, peak accent and/or are grouped by inserting a silent pause. We also observed that recall of items marked by specific contours at positions 3, 6 and 9 is particularly enhanced at these positions, whereas in sequences also containing intonational cues to items in penultimate position (2, 5 and 8) recall of those items is not equally improved. Therefore, it appears that in serial recall of spoken sequences, even when a large number of specific intonational cues to serial positions are available, listeners can use only a selection of them.

Index Terms: list intonation, serial recall, working memory

1. Introduction

In prosody research, the role of intonation in signalling discourse structure is widely acknowledged, as it cues hierarchical relationships among phrases within discourse units [1]. Specifically in lists, intonation can convey information about how they are structured into groups of items, as well as about specific item positions within a group. In this paper we explore the effects of the use of such intonational information on working memory, in particular in a serial recall task. This task consists in recalling lists of digits by following their strict serial order of presentation, and it is typically used as a test in psychology for assessing individual's short-term memory span [2]. In this research field, there is quite a large body of literature attesting the relationship between verbal serial recall and prosody, mainly consisting in the observation that lists of spoken items are better recalled when they are presented in groups (so called "grouping effect", see [3], [4], [5] among others). As to the specific role of intonation in enhancing serial recall, [6] and [7] provided evidence that it is limited to the grouping effect, i.e. it is equivalent to that triggered by pause insertion. However, in these studies results might be affected by the scarce control over the most suitable "position-informative" intonation patterns available for use in creating sequence stimuli. We hypothesise that if sequences to be recalled are produced by suitable intonation contours cueing specific positional information of items, listeners can fruitfully make use of them in terms of serial recall enhancement. In previous production and perception studies on Bari Italian, it has been

shown that a rich inventory of tunes is available to speakers for signalling hierarchical relationships within discourse units at various levels ([8], [9], [10], see comparable strategies in Dutch [11] [12]), as well as specifically for cueing positional information in sequences. The most typical contours are:

- a rising contour, signalling that the list has not yet been completed ("non-final" contour, L* L-H%);
- a high-rising contour, marking the penultimate item, i.e. signalling that the end of the list is approaching ("pre-final" contour, H* H-H%);
- a falling contour, marking the last item, i.e. cueing the end of the sequence ("final" contour, H+L* L-L%).

The aim of this study is to explore the effect of positional information cued by those tunes in a serial recall task involving Bari Italian listeners.

2. Methodology

We identified two intonational patterns (we called 'Intonation contour A' and 'Intonation contour B') characterised by F0 shapes conveying hierarchical organisation of groups within a sequence, as well as positional information of items across and within groups. In a nine-digit sequence, we determined:

- 'Intonation Contour A', consisting of the "non-final" rising contour at positions 3 and 6, and a low-falling ("final") contour at position 9. Items at initial- and within-group positions (positions 1, 2, 4, 5, 7, 8) all have a peak accent, taken to be the neutral unmarked pattern. A scheme of intonation contour A is shown in Figure 1;
- 'Intonation Contour B', sharing the same intonational patterns of 'Intonation Contour A', except for a) a steep rising pitch accent (followed by a mid-fall) at positions 2 and 5, which pre-signals the end of the first and the second groups, i.e. penultimate position in the two non-final groups within the sequence; and b) a high-rising ("pre-final") contour at position 8 pre-signalling both the end of the third group *and* the end of the whole sequence. Intonation Contour B is schematised in Figure 2.

These two experimental conditions were compared with two further ones, namely:

- 'Grouped by Pause' sequences, where all digits have a peak contour, and sequences are grouped by inserting a pause at the end of each three-digit (sub)sequence, as schematised in Figure 3;
- 'Ungrouped' (Control) sequences, sharing the same intonation of the 'Grouped by Pause', but without pause grouping, as schematised in Figure 4.

We hypothesise that serial recall performance would be:

- 1) better in both Intonation Contours A and B and the 'Grouped by Pause' conditions than in the 'Ungrouped' (Control) condition (due to the grouping effect);
- 2) better in both 'Contour A' and 'Contour B' than in the 'Grouped by Pause' condition, because of the absence of intonational marking of item position in the latter condition. In

particular, at least items in positions 3 (last item in the first serial group), 6 (last item in the second serial group), and 9 (last item in the third serial group *and* in the whole sequence) should benefit in terms of recall enhancement;

3) better in ‘Intonation Contour B’ than ‘Intonation Contour A’ because of the enhanced hierarchical and positional information conveyed by intonation in certain positions in Contour B, namely: digits at positions 2 (“pre-final” contour= item at mid position in the first group), 5 (“pre-final” contour=item at mid position in the medial group), and 8 (“pre-final” contour=item at mid position in the last group *and* penultimate item in the whole sequence).

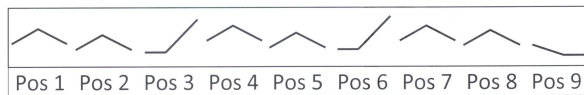


Figure 1: Schematisation of sequence stimuli as realised according to the ‘Intonation Contour A’ condition.

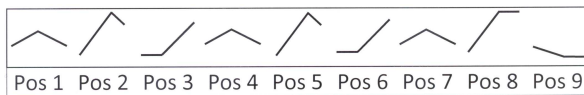


Figure 2: Schematisation of sequence stimuli as realised according to the ‘Intonation Contour B’ condition.

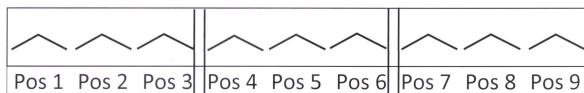


Figure 3: Schematisation of sequence stimuli as realised according to the ‘Grouped by Pause’ condition, that is a neutral, peak contour on each item, and item-grouping realised by pause insertion.

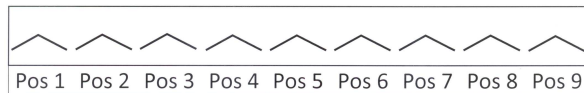


Figure 4: Schematisation of sequence stimuli as realised according to the ‘Ungrouped’ (control) condition, that is a neutral, peak contour on each item, and no item-grouping.

2.1. Stimuli

In order to produce sequences according to the conditions above, three types of stimuli for each digit (1-9) were created:

- type (a), where each digit in the sequence was realised with the unmarked, neutral F0 peak, as described above;
- type (b), where digit sequences were realised with Intonation contour A;
- type (c), where digit sequences were produced with Intonation contour B.

All series for each of the digits were produced by a trained native speaker of Bari Italian (author MS) in the same recording session. Therefore, nine sequences with the same digit (one for each digit) were produced with Intonation Contour A, nine with Intonation Contour B, and nine by realising each item with a neutral, “citation-form” intonation. In this way, all

intonational realisations in each position (first, second, third, fourth, etc.) within each contour type were available for each digit. They were saved as individual audio files, and used as “building blocks” for creating all the nine-digit spoken sequences under the four conditions. Stimuli were created by concatenating the individual audio files into nine-digit sequences. In a post-editing step, care was taken that speech signal amplitude was homogeneous in all sequences. Spoken digit realisations of type (a) were used for creating sequences for the conditions ‘Ungrouped’, and ‘Grouped by Pause’, in the latter case by inserting a 310 ms silence after digits in positions 3 and 6. Spoken digit renditions of types (b) and (c) were used for creating sequences for ‘Intonation contour A’ and ‘Intonation contour B’, respectively. An example sequence for each of the experimental conditions is shown in Figures 5-8. For example, the stimulus for the ‘Contour A’ condition shown in Figure 7 (sequence: four-seven-three six-one-eight two-five-nine) was realised by concatenating the spoken digit “four” as realised in the first position of the whole ‘digit four’ sequence with Contour A, with the spoken digit “seven” as realised in the second position of the whole ‘digit seven’ sequence with contour A, and so on.

We created 68 nine-digit lists from pseudo-random permutation of the 1-9 digits, avoiding two adjacent digits in ascending or descending order, and making sure that a digit did not appear in the same position in consecutive lists. The concatenated nine-digit sequences were created on the basis of these lists, the duration of each sequence averaging 6.4 sec. We produced 17 stimuli for each of the four conditions, for a total amount of 68 stimuli (including 8 to be used for the training session, 2 per condition). All steps for the preparation of stimuli were carried out using Praat software tool for speech analysis [13].

2.2. Subjects

Twenty-nine informants (23 females and 6 males) took part in the experimental sessions. They were aged 20-45 (average =22.4), and reported no speech or hearing deficits. They were students of Psychology at the University of Bari, all born and living in the Bari dialectal area. None of them had a background in linguistics or prosody. They were given one exam credit as a reward for participating in the experiment.

Before starting the task, subjects were tested as to their short-term memory span by means of the standard Digit Span (DS) test of WAIS-R [14], which resulted homogeneous across groups (minimum DS=5).

2.3. Procedure

Participants were tested individually in a quiet laboratory, sitting in front of a computer and wearing a headset with headphones and microphone. They were instructed to listen to each sequence and recall all the nine digits orally by strictly following their order of presentation. Spoken responses were directly recorded to disk. Each list was preceded by a warning tone and 500 ms silence, and after each spoken response subjects could proceed with the next sequence by pressing the spacebar. They were allowed to pause whenever they wanted during the session, and they were encouraged to take a break after every block of 15 stimuli. Participants were asked to recall a total of 60 sequences (15 for each condition), preceded by a short (8 stimuli) training session. The order of presentation was balanced across the subjects. The whole session (i.e. including the DS test) lasted approximately 40 min for each

informant. The experiment was implemented and run using SuperLab 2.0.

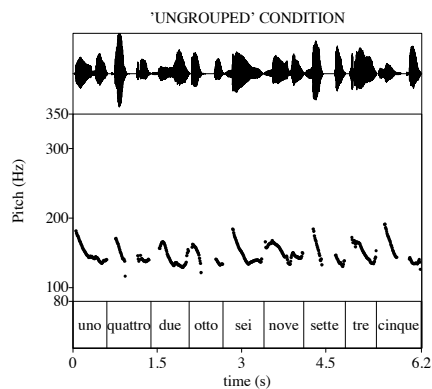


Figure 5: Speech waveform and F0 contour of one of the stimuli for the 'Ungrouped' (Control) condition.

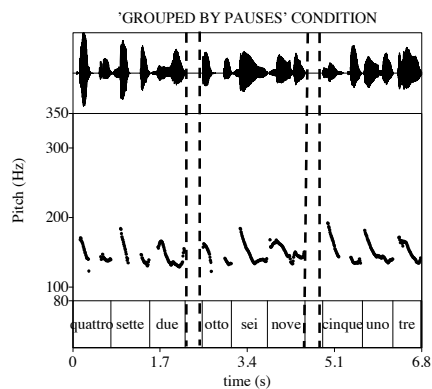


Figure 6: Speech waveform and F0 contour of one of the stimuli for the 'Grouped by Pause' condition. Vertical broken lines mark silent intervals (pauses) between groups.

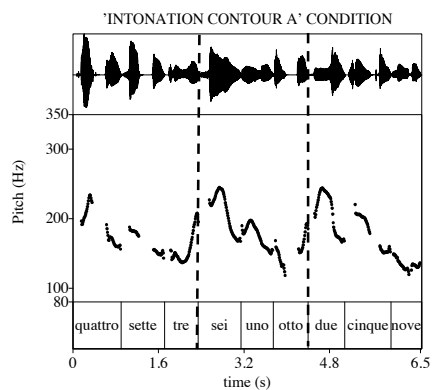


Figure 7: Speech waveform and F0 contour of one of the stimuli for the 'Intonation contour A' condition. Vertical broken lines mark the right edge of each group (intonational phrase)

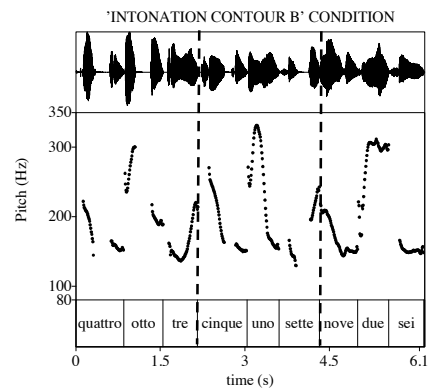


Figure 8: Speech waveform and F0 contour of one of the stimuli for the 'Intonation contour B' condition. Vertical broken lines mark the right edge of each group (intonational phrase).

3. Results and Discussion

A mixed factors general linear model was carried out, with: 1) Condition (4 levels: Ungrouped, Grouped by Pause, Intonation Contour A, Intonation Contour B), 2) Serial Group Within the Sequence (3 levels: first, second, third), 3) Within-Group Position (3 levels: first, second, third), as factors.

Results (Figures 9-12) show a very large effect of Condition: $F(3; 84) = 26.42; p < 0.001$, in that sequences in the 'Ungrouped' (Control) condition are recalled significantly worse than those in the three remaining conditions. This confirms the general findings on serial memory of verbal sequences that they are recalled better when they are grouped prosodically ("grouping effect"). Most interestingly, lists produced with both Contours A and B show statistically better recall performance with respect to those produced by inserting silent pauses between serial groups (Fisher LSD post hoc test, $p < 0.05$). This confirms our prediction that intonation plays a specific role in serial recall enhancement beyond the grouping effect. On the other hand, our hypothesis that Contour B sequences would perform better than Contour A sequences is not confirmed. We also found a significant second order interaction effect between Condition, Serial Group and Within-Group Position: $F(12; 336) = 3.06; p < 0.001$; again, Contour A and Contour B conditions showed better recall performance than sequences Grouped by Pause, but without any difference between Contour A and Contour B sequences. We then looked at recall performance of specific intonationally-marked positions in the sequences, namely positions 3, 6 ("non-final" contour in both Contour A and Contour B conditions), positions 2, 5, 8 ("pre-final" contour in Contour B condition), and position 9 ("final" contour in both Contour A and Contour B conditions). We performed Planned Comparisons between these positions in the relevant conditions, namely:

- position 3 in Contours A/B vs. Grouped by Pause conditions: recall performance was significantly better in both Contour A and Contour B ($p < 0.01$);
- position 6 in Contours A/B vs Grouped by Pause conditions: recall performance was significantly better in Contour B ($p < 0.05$), but not in Contour A (only approaching statistical significance, $p = 0.06$);

- position 9 in Contours A/B vs Grouped by Pause conditions: again, recall performance was significantly better in both Contours A and B ($p < 0.01$);
- positions 2, 5, 8 in Contour A vs Contour B conditions: recall performance did not show any statistical difference.

These outcomes indicate that serial recall of items marked by “non-final” contours (i.e. those marking positions corresponding to the end of a serial group) and “final” (marking the end of a sequence) is particularly enhanced. Interestingly, the effect of positional information cued by a “final” F0 contour (position 9) is so strong that it significantly increases the recency effect. On the other hand, listeners did not make use of “pre-final” intonational information during the recalling phase, which is the reason why overall recall performance of Contour B sequences was not significantly better than those with Contour A, contrary to our hypothesis.

Therefore, it appears that there is a limit to the number of intonational cues to positions which listeners can use, and that those marking the end of a serial group (at positions 3 and 6) or the end of the sequence (at position 9) are the most likely to be used.

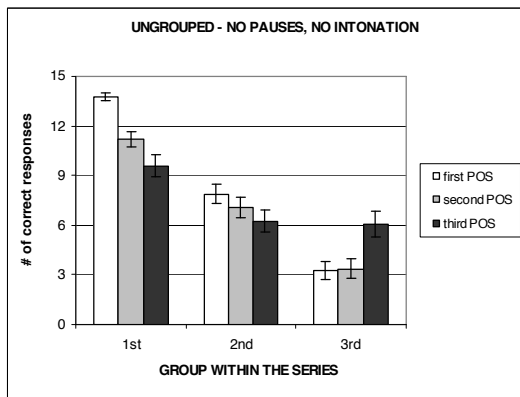


Figure 9: Correct recall (mean values) across the 3 serial groups within a sequence, and the position within each serial group (first POS, second POS, third POS), for the ‘Ungrouped’ (Control) condition.

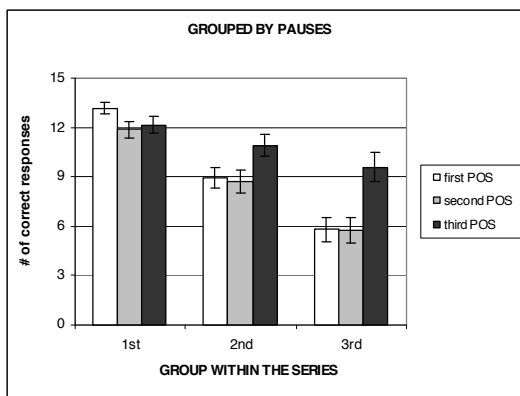


Figure 10: Correct recall (mean values) across the 3 serial groups within a sequence, and the position within each serial group (first POS, second POS, third POS), for the ‘Grouped by Pause’ condition.

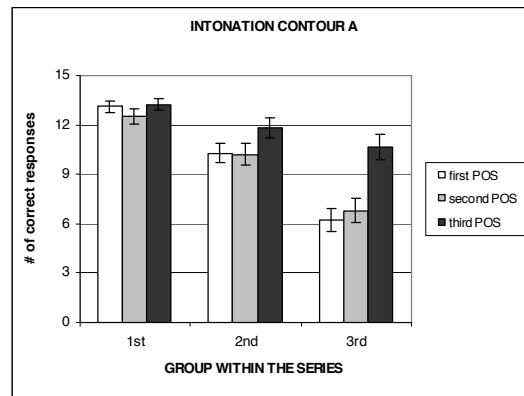


Figure 11: Correct recall (mean values) across the 3 serial groups within a sequence, and the position within each serial group (first POS, second POS, third POS), for the ‘Intonation Contour A’ condition.

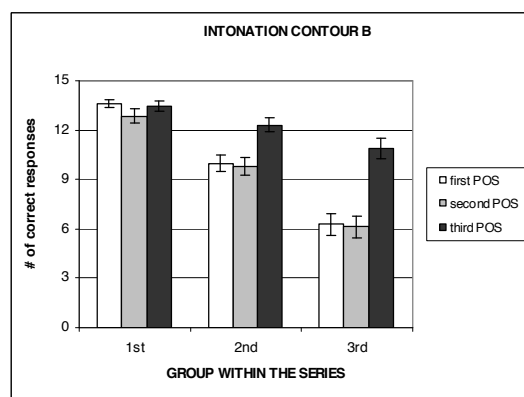


Figure 12: Correct recall (mean values) across the 3 serial groups within a sequence, and the position within each serial group (first POS, second POS, third POS), for the ‘Intonation Contour B’ condition.

4. Conclusions

In this paper, we explored the effects of positional information conveyed by intonation on serial memory. In a serial recall task, we observed that when listeners are presented nine-digit lists with F0 contours marking the hierarchical organisation of items within the sequence, their recall performance is significantly better than in cases where such intonational information is absent. However, it appears that intonational enhancement of serial recall has its limits, since we also observed that even when a large number of specific intonational cues to serial position are available, listeners are only able to make use of a selection of them. Such limitations might be due to cognitive processing needs imposed by working memory.

5. Acknowledgements

We would like to thank Ralf Rummer for helpful discussion and advice, and our student assistants in Bari Mirco Lacalandra and Gabriella Monticelli for help collecting the experimental data.

6. References

- [1] Hirschberg, J. and Pierrehumbert, J. "The intonational structuring of discourse", Proceedings of the 24th Annual Meeting of the Association of Computational Linguistics, 136-144, 1996.
- [2] Baddeley A., Eysenck M.W., Anderson, M. C., "Memory", Psychology Press, 2009.
- [3] Frankish, C., "Modality-specific grouping effect in short-term memory", *Journal of Memory and Language*, 24: 200-209, 1985.
- [4] Cowan, N. Saults, J. S., Elliott, E. M., and Moreno, M. V., "Deconfounding serial recall", *Journal of Memory and Language* 46:153-177, 2002.
- [5] Reeves, C., Schmauder A. R., and Morris, R. K. "Stress grouping improves performance on an immediate serial list recall task", *Journal of Experimental Psychology: Learning, Memory and Cognition*, 26 (6): 1638-1654, 2000.
- [6] Frankish, C. "Intonation and auditory grouping in immediate serial recall", *Applied Cognitive Psychology*, 9: 5-22, 1995.
- [7] Saito, S., "Effects of articulatory suppression on immediate serial recall of temporarily grouped and intonated lists", *Psychologia*, 41: 95-101, 1998.
- [8] Savino, M., "Non-finality and pre-finality in Bari Italian intonation: a preliminary account", Proceedings of the VII European Conference on Speech Communication and Technology, 2: 939-942, 2001.
- [9] Savino, M., "Intonational cues to discourse structure in a variety of Italian", in P. Gilles and J. Peters [Eds], *Regional variation in intonation*, 161-187, Niemeyer, 2004.
- [10] Savino M., Grice, M., Gili Fivela, B. Marotta, G., "Intonational cues to discourse structure in Bari and Pisa Italian: perceptual evidence", Proceedings of Speech Prosody (on CD-ROM), 2006.
- [11] Geluykens, R., Swerts, M., "Prosodic cues to discourse boundaries in experimental dialogues", *Speech Communication*, 15: 69-77, 1994.
- [12] Swerts, M., Collier, R., Terken, J., "Prosodic predictors of discourse finality in spontaneous monologues", *Speech Communication*, 15: 79-90, 1994.
- [13] Boersma, P. and Weenink, D., "Praat, a system for doing phonetics by computer", *Glott International*, 5(9/10): 131-151, 2001.
- [14] Wechsler, D., "Manual for WAIS-R" The Psychological Corporation, 1987.