Title
Intonation and positional effects in spoken serial recall

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN
1069-7977

Authors
Savino, Michelina
Bosco, Andrea
Grice, Martine

Publication Date
2013

Peer reviewed

Intonation and Positional Effects in Spoken Serial Recall

Michelina Savino (michelina.savino@uniba.it)
Dipartimento di Scienze della Formazione, Psicologia, Comunicazione, Università di Bari, piazza Umberto I, 1
70121 Bari, ITALY

Andrea Bosco (andrea.bosco@uniba.it)
Dipartimento di Scienze della Formazione, Psicologia, Comunicazione, Università di Bari, piazza Umberto I, 1
70121 Bari, ITALY

Martine Grice (martine.grice@uni-koeln.de)
IfL-Phonetik, Universität zu Köln, Herbert-Lewin-Strasse 6
50931Cologne, GERMANY

Abstract

Past studies have indicated that intonation, in the sense of fundamental frequency modulation, can only enhance serial recall to the extent that it can induce a grouping effect, something that can also be induced by a simple insertion of pauses. However, in a study of spoken serial recall of nine-digit lists, we are able to show that recall is significantly better when sequences of digits are marked by specific intonation contours than when they are simply grouped by silent pauses in the signal. Thus, we found that intonation plays a role during the encoding phase, whereby items in group-final positions draw particular benefit from intonation. However, intonation does not appear to play the same role in the retrieval phase, since when subjects are instructed to imitate intonation during recall, performance shows mixed effects.

Keywords: serial recall; intonation; grouping effect; short-term memory

Introduction

In field research, there is a general consensus that serial recall (short-term memory in general) and prosody are closely related. Well-documented evidence of such a relationship is the grouping effect, that is, the enhanced recall of items in a list when they are presented in groups (for example, Reeves et al 2000). The grouping effect is stronger for auditory stimuli (Cowan et al 2002, Frankish 1985), as prosody plays an important role in this grouping, or patterning. Past and more recent research has aimed to ascertain the nature of these groups. In their seminal work on the auditory grouping effect, Frankish (1995) and Saito (1998) provided evidence that it can be obtained by temporal organisation of speech stimuli realised by pause insertion between groups as well as by superimposing a "natural" intonational pattern, or by manipulating pitch levels on groups of items (Frankish 1995). On the other hand, more recent studies have shown that the grouping effect in serial recall reflects rhythmic groups, referred to as the stress grouping effect, rather than intonational phrases (Reeves et al 2000, Boucher 2006, Gilbert & Boucher 2007). It is argued that these groups correspond to the segmentation units (chunks) listeners use in spoken language perception (Gilbert et al 2011).

Previous research dealing specifically with the role of intonation in improving serial recall suggests that it is relevant to the extent that it can induce a similar grouping effect to that obtained by pause insertion (Frankish 1995, Saito 1998). However, a number of potentially relevant aspects of intonation deserve further exploration. For instance, in the previous studies discussed above, the superimposition of a fundamental frequency (F0) contour[1] on the whole sequence, or F0 manipulation on groups of items within a sequence were carried out with little control over specific tunes and their associated meanings. This is particularly relevant, since the role of intonation in signalling discourse structure is widely acknowledged, as it cues hierarchical relationships among phrases within a discourse unit (Hirschberg & Pierrehumbert 1986). Moreover, in sequences intonation can convey information about the hierarchical structure (groups) as well as about specific positions within a group. In Italian (in particular the variety of Bari), a rich inventory of tunes is available for marking those kinds of hierarchical relationships in a sequence (Savino 2001; 2004), among which the most typical are:

- The "continuation rise" contour, a gradually rising F0 movement from the nuclear syllable up to the end of the phrase. It signals that the list has not been completed yet, and that more items are to come ("non-finality" contour);
- A high rising contour, where the rise in F0 starts before the nuclear syllable and continues rising up to the end of the phrase. It conveys the information that the current item is the penultimate in a sequence, i.e. that the end of the list is approaching ("pre-finality" contour");
- A falling contour, involving a gradual fall from the nuclear syllable until the end of the phrase. This contour marks the end of a sequence ("finality" contour).

Our aim here is to verify whether the use of specific tunes conveying such hierarchical relationships and positional

[1] The F0 contour corresponds roughly to what is perceived as the pitch, or melody.

information within a sequence could improve immediate serial recall performance of auditory stimuli in (Bari) Italian listeners.
Another positive effect in serial recall performance reported in the literature is the salience effect. It "occurs when an item that is conspicuous on some perceptual dimension is recalled better than other items in the same ensemble in learning and memory tasks" (Reeves et al 2000: 1639). In this paper we also explore the role of pitch prominence in within-group (medial) positions in the sequence as another potential factor in serial recall enhancement.
Moreover, we evaluate whether intonation plays a specific role not only in encoding but also in retrieval of phrases, i.e. whether performance is improved when listeners imitate the tune of the input stimuli during recall.

Method and Materials

Intonational Patterns

In order to investigate the role of specific tunes in serial recall, we identified two intonational patterns (we called 'intonation contour A' and 'intonation contour B') characterised by F0 shapes conveying hierarchical organisation of groups, as well as positional information of items across and within groups. In a nine-digit sequence, we determined:
- 'Intonation contour A', consisting of the "continuation rise" ("non-final") contour at positions 3 and 6, and a low-falling ("final") contour at position 9. F0 shapes of items at initial and within-group positions (positions 2, 5, 7, 8) are not positionally marked by intonation, as they are characterised by a peak accent (taken to be the neutral unmarked pattern).
A schematisation of intonation contour A is shown in Figure 1;
- 'Intonation contour B', sharing the same intonational patterns of 'intonation contour A', except for a steep rising pitch accent at positions 2 and 5. It can pre-signal the end of the first and the second groups, i.e. the two non-final groups within the signal. Also, because of the steep rising accentual movement, these digits sound perceptually more salient than the corresponding positions 2 and 5 in intonation contour A, where they are marked by a (neutral) peak accent instead. Another feature of contour B is 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.
A schematisation of 'intonation contour B' is given in Figure 2.
These two experimental conditions were compared with two further ones, namely:
- 'Grouped by pauses' sequences, where all digits have a peak contour, and sequences are temporally grouped into three by inserting a pause at the end of each group;
- 'Ungrouped' sequences, sharing the same intonation of the 'grouped by pauses' stimuli above, but without pause grouping.
We hypothesise that serial recall performance would be
(H1) better in both intonation contours A and B and the 'grouped by pauses' conditions than in the 'ungrouped' (control) condition, due to the grouping effect;
(H2) better in both 'contour A' and 'contour B' than in the 'grouped by pauses' condition, because of the absence of intonational marking of item position in the latter condition. In particular, items in positions 3 (non-final contour=last item in the first group), 6 (non-final contour=last item in the second group), and 9 (final contour=last item in the third group *and* in the whole sequence) should benefit in terms of recall enhancement.
(H3) better with 'intonation contour B' than 'intonation contour A' because of the enhanced hierarchical and positional information conveyed by intonation in certain positions, 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 = penultimate item in the whole sequence) This should result in an overall better recall performance of 'intonation contour B' than of 'intonation contour A' sequences.
Moreover, we hypothesise that recall performance in both intonation conditions
(H4) would be enhanced for subjects who are instructed to imitate the intonation produced in the stimuli during the recall task. This would provide some evidence that intonation plays a relevant role not only in encoding but also in retrieval of verbal material.
Our hypotheses can be summarized as follows:
Contour B > *Contour A* > *Grouped by pauses* > *Ungrouped* on the one hand, and *Imitation* > *No imitation* on the other.

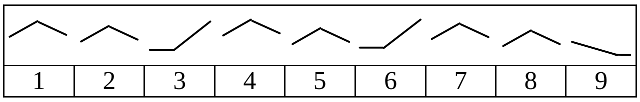

Figure 1: Schematisation of "intonation contour A"

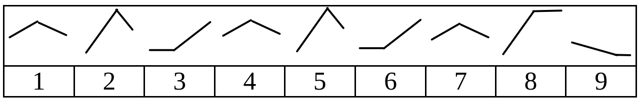

Figure 2: Schematisation of "intonation contour B"

Stimuli

The sequences were created according to the four conditions ('ungrouped', 'grouped with pauses', 'intonation contour A', 'intonation contour B'), three types of spoken stimuli for each digit were created:
Type (a), where each digit was realised with a neutral F0 peak, as, for example, in digit position 1 in contours A and B (see schematisation in Figures 1-2).
Type (b), where sequences were realised with intonation contour A described above (see Figure 1)
Type (c), where sequences were produced with intonation contour B described above (see Figure 2)

All series for each of the digits were produced by a trained native speaker of Bari Italian (author MS) in the same recording session. In this way, for each digit all intonational realisations in each position (first, second, third, etc.) within each sequence type were available. They were saved as individual audio files, and used as "building blocks" for creating all the nine-digit spoken sequences under the 4 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.
This methodology enabled us to create stimuli which sound more natural than those produced by means of a speech synthesiser, as in Frankish (1995).
Spoken digit realisations of type (a) were used for creating sequences for the conditions 'ungrouped', and 'grouped by pauses', in the latter case by inserting a 310 ms silence after digits in positions 3 and 6. Spoken digit renditions of the types (b) and (c) were used for creating sequences under the conditions 'intonation contour A' and 'intonation contour B', respectively.
We created 68 nine-digit lists from pseudo-random permutation of the 1-9 digits, in a way to avoid the presence of two adjacent digits in ascending or descending order, and to make sure that a digit would 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 only, 2 per condition).
All steps for the preparation of stimuli were carried out using the Praat software tool for speech analysis (Boersma & Weenink 2001). Examples of a sequence (speech waveform and F0 contour) for each of the four conditions are shown in Figures 3-6.

Informants

Fifty-six subjects participated in the experiment. They were undergraduate and graduate students of Psychology at the University of Bari (average age: 23.6), with no reported speech or hearing deficits. They were all native speakers of Italian, born and living in the Bari geolinguistic area. None of them had a background in phonetics or prosody.

Procedure

Participants was tested individually in a laboratory. They were asked to listen to each sequence and recall all nine digits orally by strictly observing their serial order. Half of them were also instructed to imitate the intonation of the sequence during the recall.
Subjects were seated in front of a computer and wore a headset with headphones and microphone. Each list was preceded by a warning tone and 500 ms silence. Spoken responses were recorded directly and subjects proceeded to the next sequence by pressing the spacebar. They were also free to take a break any time they needed during the session. A break was suggested after every block of 15 stimuli.
For each subject, the task involved recalling 60 stimuli (preceded by a short training session) i.e. 15 stimuli for each of the four conditions. The order of presentation was balanced across the subjects.
Before starting the experimental trial, participants were tested for their digit span of the WAIS-R (Wechsler 1987). This step was carried out in order to ascertain that the digit span of subjects was homogeneous across groups.
The average total duration of sessions (including the digit span test) was around 40 min. Sessions were implemented and run using SuperLab 2.0.

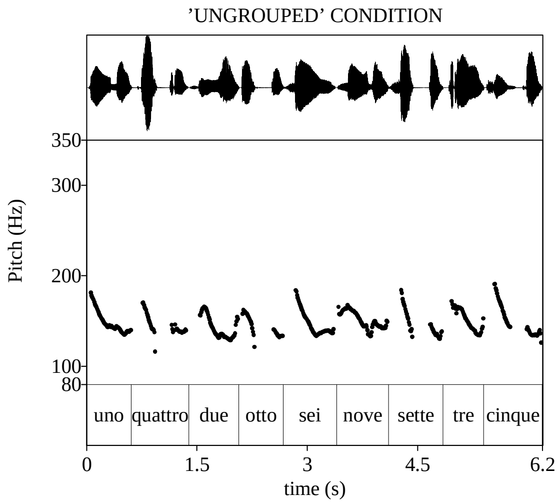

Figure 3: Speech waveform and F0 contour of one of the stimuli for the 'ungrouped' (control) condition.

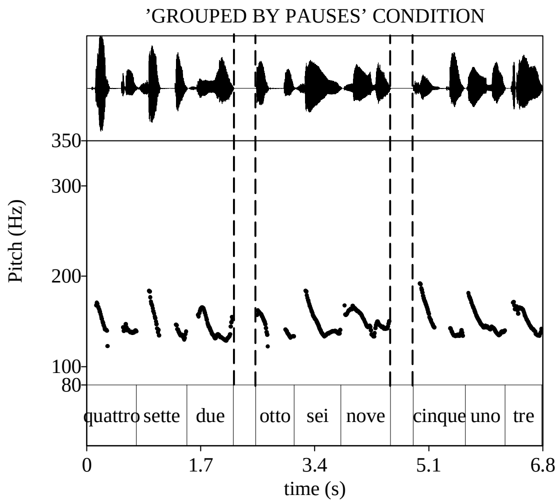

Figure 4: Speech waveform and F0 contour of one of the stimuli for the 'grouped by pauses' condition. Vertical broken lines mark silence intervals (pauses) between groups.

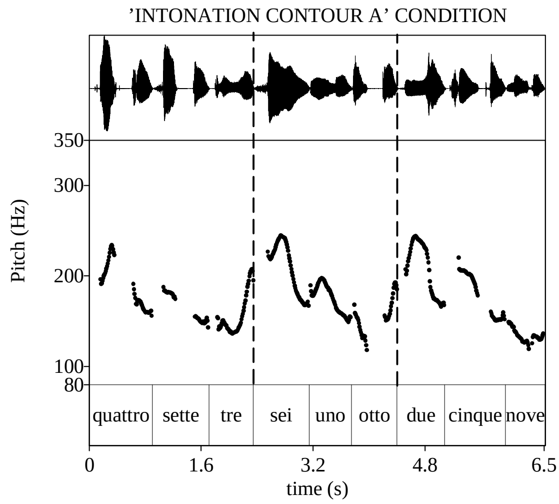

Figure 5: Speech waveform and F0 contour of one of the stimuli for the 'intonation contour A' condition. Vertical broken lines mark the right edge of each group (intonational phrase)

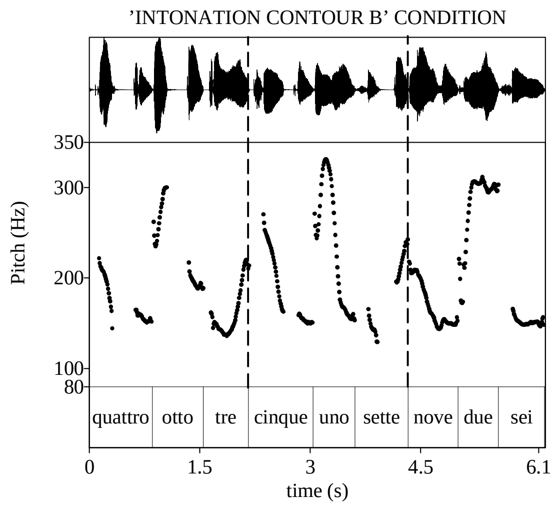

Figure 6: Speech waveform and F0 contour of one of the stimuli for the 'intonation contour B' condition. Vertical broken lines mark the right edge of each group (intonational phrase)

Results and Discussion

A mixed factors general linear model was performed with:
1) condition (within subject), 4 levels (ungrouped, grouped by pauses, intonation A, intonation B)
2) imitation of intonation (between subjects), 2 levels (yes, no)
3) serial group within the sequence (within subjects), 3 levels (first group, second group, third group)
4) within group position (within subjects), 3 levels (first position, second position, third position)
as factors.

First of all, a large effect of condition was found: $F(3; 159) = 52.75$; $p < 0.001$; partial eta square = 0.5, performance percentages for each condition: ungrouped=52%, grouped by pauses=64%, intonation contour A=70%, intonation contour B=69%). These results are in line with our prediction in (H1), and confirm the typical grouping effect in immediate serial recall of prosodically manipulated auditory stimuli (Frankish 1995, Saito 1998).

Importantly, our hypothesis in (H2) is also verified, since results show a statistically significantly better recall performance for sequences with intonation contours A and B than for those grouped simply by pauses (post-hoc $p<0.01$). These results point to a specific role of intonation in serial recall that goes beyond the grouping effect induced by pauses, counter to the claims in previous studies.

We also found a significant second order interaction between condition, serial group and within-group position: $F(12 ; 636) = 2.35$; $p < 0.01$; partial eta square = 0.04, with a moderate effect. Results (Figures 7-10) relating to positions in the first and second serial groups show a better performance for sequences realised with contours A and B, with respect to those grouped by pauses. The effect is particularly relevant for positions 3 and 6 (i.e. the last position in the first and second groups). This can be seen by comparing the "descending" shape of the correct response bars for groups 1 and 2 in the ungrouped condition (Figure 7), with the "dipped" shape for the same group in the remaining 3 conditions (Figures 8-10). The reverse tendency is evident for the third group, with an "ascending" shape in all conditions, pointing to an enhancement of the recency effect in all conditions.

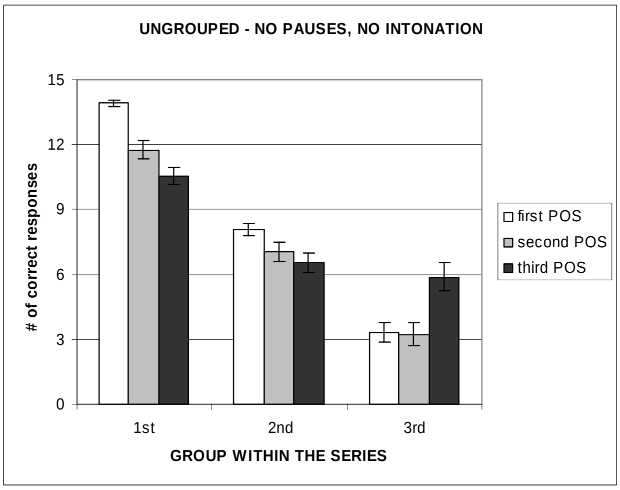

Figure 7: Number of correct recalls (mean values) across the 3 serial groups in the sequence and the positions within each group (first POS, second POS, third POS), for the 'ungrouped' (control) condition.

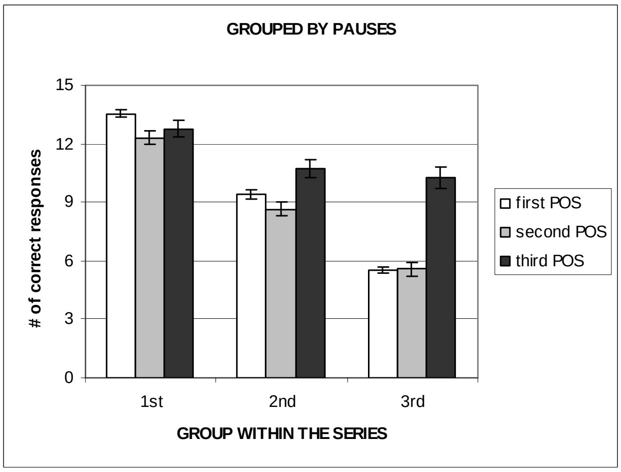

Figure 8: Number of correct recalls (mean values) across the 3 serial groups in a sequence and the positions within each group (first POS, second POS, third POS), for the ‘grouped by pauses’ condition.

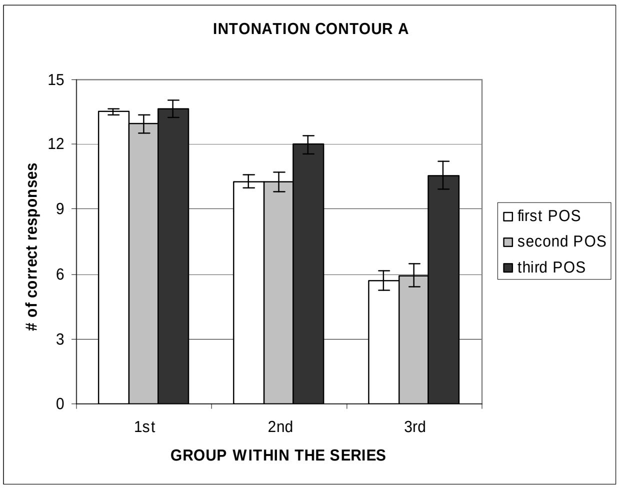

Figure 9: Number of correct recalls (mean values) across the 3 serial groups in a sequence and the positions within each group (first POS, second POS, third POS), for the ‘intonation contour A’ condition

Hypothesis (H3) is not confirmed by our results, as we did not obtain the expected specific effect in contour B for the second (pre-final) position within the serial groups (i.e. positions 2, 5 and 8 within the sequence, where 8 is also the penultimate item position in the whole sequence).
For position 5, the salience potentially conveyed by the steep rising pitch accent (compared to the “neutral” pitch peak in the same position in contour A) does not appear to improve recall of the medial item in a sequence.
With regard to Hypothesis H4, we did not find a main effect of the imitation of intonation, even though this factor interacts with a moderate effect with the serial groups: $F(2; 106) = 5.82$; $p < 0.01$; partial eta square = 0.1). Diagrams in Figure 11 show that imitation of intonation improves the recall of the first group within a sequence, whereas performance decreases for the second and third serial groups. Thus, the imitation of intonation does not appear to facilitate recall, except for the digits in the first group, when the memory traces are stronger (as confirmed by low recency effects in our data).

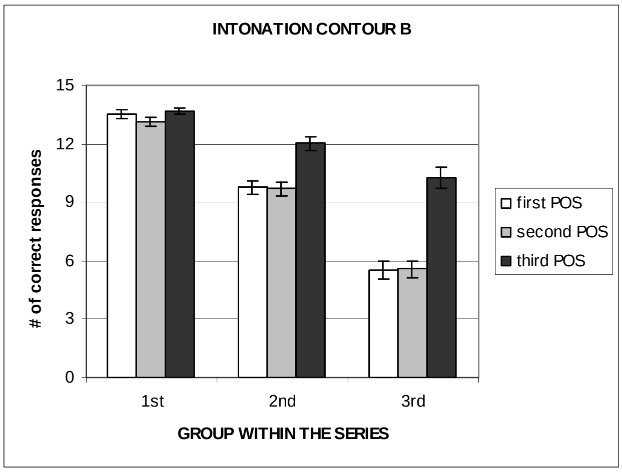

Figure 10: Number of correct recalls (mean values) across the 3 serial groups in a sequence and the positions within each group (first POS, second POS, third POS), for the ‘intonation contour B’ condition

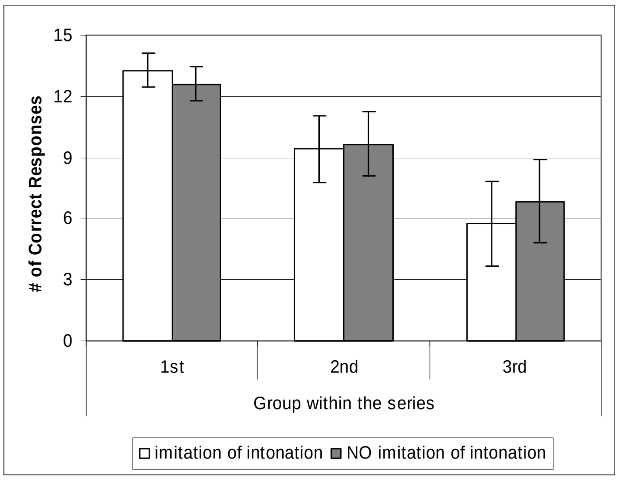

Figure 11: Number of correct recalls across the three serial groups as a function of instructions to imitate the intonation.

Conclusions

The aim of our study was to ascertain whether the use of intonation patterns conveying hierarchical relationships among and within groups in a sequence could improve performance in an immediate serial recall task. Output modality is oral channel, as in serial recall output modality can have a considerable influence in task performance (Penney 1979). Our results show that oral recall of spoken lists is better when auditory stimuli are intonationally marked by specific tunes than when they are simply temporally grouped by pauses. Some positions particularly benefitted from intonation: a rising non-final contour marking serial positions 3 and 6 is more effective than pauses.

On the other hand, we did not find clear evidence that a pre-final contour in positions 2 and 8 (marking penultimate position within the first group, and penultimate position within the whole sequence, respectively) can enhance recall of items in these positions, also possibly because of the masking effect of primacy and recency in adjacent positions. It also appears that the item in medial position within the whole sequence (position 5) cannot be better recalled when made perceptually salient by pitch prominence.

These results indicate that intonation can be attributed a role in the enhancement of serial recall that is not simply equivalent to the temporal structuring induced by pauses, as has so far been claimed. They also provide further support to the "language-oriented" view of Short-Term Memory (STM) as opposed to models of STM that propose a specialised memory system that is hierarchically organised (Farrell & Lelièvre 2012).

We also found that particular intonation patterns can improve recall only in the input modality, i.e. when it is presented in auditory stimuli, but not in the output modality, i.e. when subjects are requested to imitate the intonation of the auditory stimuli whilst recalling. In the latter case, imitation cannot be construed as a supporting strategy for the recalling task. Instead it interfered with the task, resulting in a decrease in performance in the last part of the sequence, as memory traces become weaker.

Such result that imitation of the intonation does not enhance serial recall points to supporting the "language-oriented" models of Short-Term Memory.

Acknowledgments

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

References

Boersma, P. & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5(9/10)*, 131-151.

Boucher, V. J. (2006). On the function of stress rhythms in speech: evidence of a link with grouping effects on serial memory. *Language and Speech 49(4)*, 495-519.

Cowan, N. Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language, 46*, 153-177.

Farrell S. & Lelièvre A. (2012). The dynamics of access to groups in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 38(6)*, 1659-1674.

Frankish, C. (1985). Modality-specific grouping effect in short-term memory. *Journal of Memory and Language*, *24*, 200-209.

Frankish, C. (1995). Intonation and auditory grouping in immediate serial recall. *Applied Cognitive Psychology*, *9*, 5-22.

Gilbert, A. C. & Boucher, V. (2007). What do listeners attend to in hearing prosodic structures? Investigating the human speech-parser using short-term recall. *Proceedings of Interspeech* 2007 (pp.430-433).

Gilbert, A. C., Boucher, V. & Jemel, B. (2011). The role of rhythmic chunking in speech: synthesis of findings and evidence from statistical learning. *Proceedings of the XVII International Congress of Phonetic Sciences* (pp. 747-750).

Hirschberg, J & Pierrehumbert, J. (1996). The intonational structuring of discourse. *Proceedings of the 24th Annual Meeting of the Association of Computational Linguistics* (pp.136-144).

Penney, C. G. (1979), Interactions of suffix effects, with suffix delay and recall modality in serial recall. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 507-521.

Reeves, C., Schmauder A. R., & Morris, R. K. (2000). Stress grouping improves performance on an immediate serial list recall task. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26 (6)*, 1638-1654.

Saito, S. (1998), Effects of articulatory suppression on immediate serial recall of temporarily grouped and intonated lists. *Psychologia, 41*, 95-101.

Savino, M. (2001). Non-finality and pre-finality in Bari Italian intonation: a preliminary account. *Proceedings of the VII European Conference on Speech Communication and Technology* (vol.2, pp.939-942).

Savino, M.(2004). Intonational cues to discourse structure in a variety of Italian. In P. Gilles & J. Peters (Eds.), *Regional Variation in Intonation*, Tuebingen: Niemeyer.

Wechsler, D. (1987). *Manual for WAIS-R*. New York: The Psychological Corporation.