Consequences of Working Memory Differences and Phrasal Length on Pause Duration and Fundamental Frequency

Caterina Petrone[1*], Susanne Fuchs[1], Jelena Krivokapic[2]

[1]ZAS, Centre for General Linguistics
Schützenstr. 18, Berlin, Germany

[2]Yale University
Linguistics Department, 370 Temple St., New Haven, USA

{petrone,fuchs}@zas.gwz-berlin.de jelena.krivokapic@yale.edu

Abstract. *It has recently been suggested that speakers vary in the amount of speech planning and that the scope of planning is influenced by task- and speaker specific constraints. To test this, an experiment is presented examining the effects of linguistic structure and working memory on speech planning, as evidenced in pause duration and F0 peaks. Twenty speakers of German performed two tasks. In the first task, speakers' working memory span was evaluated. In the second, a reading task which was acoustically recorded, the influence of phrasal length on pause duration and on the utterance initial F0 peak was tested. The hypothesis is that speakers with higher WM span will show evidence for larger scopes of planning compared to speakers with low WM span, such that they will have longer pause duration and their F0 will start higher. Results show an effect of phrase length and of WM span on F0. The implications of these findings for models of speech planning are discussed.*

1. Introduction

In this paper we investigate the phonetic consequences of the relationship between linguistic (utterance length) and cognitive (working memory) factors on speech planning. It is widely agreed upon that language production proceeds in an incremental fashion, with speakers articulating some parts of the utterance while planning the upcoming ones (e.g., Levelt, 1989; Wheeldon & Lahiri, 1997). However, different units of planning have been proposed, ranging from the prosodic word to the intonational phrase (e.g., Levelt, 1989; Wheeldon & Lahiri, 1997; Krivokapić, 2007). These discrepancies have led some researchers to suggest that the scope of planning does not coincide with a fixed linguistic unit, but might be flexibly adapted by the speaker (Ferreira & Swets, 2002; Swets et al., 2007; Wagner et al., 2010). We hypothesize that such speaker-specific differences can result from differences in working memory (henceforth WM) capacities. "Working memory" refers to the memory sub-system

[*] Supported by the BMBF (Grant Nr. 01UG0711) and a grant by the DFH to the PILIOS project.

which is responsible for the active maintenance of mental representations (such as action plans or goal states) in face of ongoing processing and/or distractions (Conway et al., 2005). In particular, Swets et al. (2007) argued, based on their findings in listener comprehension of syntactically ambiguous sentences, that there is an effect of WM capacities on prosodic structure, such that readers with low WM are more likely to chunk a text into smaller prosodic phrases than readers with high WM. The link between cognitive constraints and speech planning has been more directly investigated by Wagner et al. (2010). By measuring utterance initiation times they found out that an increase of cognitive load lead speakers to narrow down the scope of planning, thus providing evidence that the scope of planning is flexible.

A number of studies has investigated phonetic parameters that reflect the relationship between planning and linguistic structure. For instance, it has been shown that the duration of the pause preceding an utterance is longer when the following utterance is longer or syntactically/prosodically more complex (Cooper & Paccia-Cooper, 1981; Zvonik & Cummins, 2003; Krivokapić, 2007). These findings are accounted for in terms of planning, since it is assumed that the speakers employ the pause before utterance onset to plan the upcoming utterance. On the other hand, it is also known that there is extensive variability in pause duration and pause placement (Goldman Eisler, 1968; Zvonik & Cummins, 2003). While these differences might stem from many factors, one possibility is that they reflect speaker-specific differences in planning.

Another phonetic parameter of planning is fundamental frequency (F0). F0 is the primary acoustic cue for intonation and the idea of "intonation planning" has often been examined on the basis of F0 "declination", i.e., the gradual F0 downtrend across the utterance. A way to test whether F0 declination is planned at a more global or a more local level is to look at the height of the first F0 peak. It is assumed that, if F0 declination is planned at a more global level, speakers will start with higher F0 peaks when producing longer utterances as a strategy to accommodate for declination in larger prosodic chunks. For instance, Prieto et al. (2006) compared the height of utterance initial F0 peaks in relation to the length of the Subject constituent in different Romance languages, with the Subject being uttered as a single prosodic constituent in Subject-Verb-Object sentences. They found that only some speakers started with higher F0 peaks when producing longer utterance, thus suggesting that planning of F0 declination is not obligatory, but an optional mechanism by the speaker. Furthermore, in a recent work on German (Fuchs et al., submitted), we investigated the effects of length on pause duration and on the utterance-initial F0 peak in Subject-Verb-Object sentences. We found that pause duration is not sensitive to length manipulations, and that the variability across speakers is quite small. On the other hand, the utterance-initial F0 peak is significantly higher in longer Subject constituents. Similarly to Prieto et al.‘s findings, some speakers showed stronger effects of Subject length on the F0 values, while others showed smaller or no effects. The reasons of such speaker-specific differences are still unknown.

The goal of the current study is to examine whether possible individual differences in planning might be accounted for in terms of differences in WM capacities. Following Swets et al. (2007), we predict that the scope of planning is narrower in speakers with low WM span than in speakers with high working memory span. To test this, we reanalyze the same acoustic dataset reported in Fuchs et al. (submitted) by linking the speaker-specific variability in pause duration and utterance-initial F0 peaks to differences in WM capacities in the same speakers. We hypothesize that the relationship between WM and planning will be more visible for utterance-initial F0 peaks, since in our data their acoustic realization is characterized by stronger speaker variability than pause duration. As a consequence, we predict that speakers with high WM span will start with higher F0 peaks regardless of phrase length, since they are able to plan larger prosodic constituents. In what follows we present the experiment examining these hypotheses.

2. Methods

2.1. Participants

Twenty native speakers of German (9 males and 11 females) participated in the experiment. Speakers performed first the WM test and then the reading task, which was recorded acoustically. This presentation order was kept fixed across speakers to avoid possible effects of tiredness on the working memory scores. The WM test and the acoustic task are detailed below (the acoustic task has also been described in Fuchs et al. (submitted), Experiment 2).

2.2. Working memory test

A modified and automated version of the WM task was employed (Conway et al., 2005), aiming at grouping the speakers according to their WM capacity (speakers with high vs. low WM capacity). The WM test was run on a UNIX machine in a quiet room at ZAS (Berlin). Participants were seated in front of a computer screen and they were asked to perform a WM reading span task, which included both a recall and a semantic verification tasks. The recall task consisted of 12 trials, each containing words to be remembered. All words were German monosyllabic, high frequent words. Single trials consisted of 2, 3, 4 or 5 to-be-remembered words. Trials of each set size were presented in random order. Each word appeared for a few seconds in isolation on the computer screen. Participants were asked to silently read the words and to remember them. At the end of each trial, they had to type the memorized words via the keyboard. Between these words, semantically complex sentences were presented. Participants were asked to perform a semantic verification task on them, i.e. they had to determine whether each sentence was semantically right or wrong. The semantic verification task was chosen to "distract" the participants from the to-be-remembered words and had the aim to avoid the adoption of memory rehearsal strategies. The duration of the visual presentation of each sentence was speaker-adjusted during a prior familiarization phase to further minimize memory rehearsal. The WM span task lasted on average 10-15 minutes for each participant. The score of the WM span task was automatically calculated in terms of partial-credit unit (PCU) scoring, by which the WM score for each trial is calculated as the proportion of words recalled in that trial (Conway et al., 2005). For instance, in a trial with 4 words a participant will score 0.25, if he/she remembers only 1 out of 4 words and 0.5 if he/she remembers 2 out of 4 words. The total proportion score for each participant was the average of the proportion scores across all trials. Participants were then split into high vs. low WM capacity sub-groups depending on whether their WM score was above or below the median WM score of the whole group. The results for the recall and semantic verification tasks were imported in R for statistical analyses.

2.3. Acoustic task

Subsequent to the WM reading span task, speakers read aloud a series of target sentences, which were constructed to test the effects of utterance length on pause duration and on the utterance-initial F0 peak. The corpus consisted of two sets of sentences composed of Subject-Verb-Prepositional phrases (S-V-P). The S constituent and P phrase were proper names modified in length. The S constituent had three length conditions (short = 4 syllables, medium = 5 syllables, long = 7 syllables) whereas the P phrase had only two (short = 2 syllable, long = 6 syllables). The two factors were crossed. The material between the S constituent and the P phrase was kept constant at 7 syllables. Each experimental condition had two sentences, one where the S constituent started with *Lena* and another which was identical except that it started with *Lilli.* A F0 peak (analyzed as a prenuclear H* accent in German) was expected to be realized on each of these two words. Moreover, the difference in stressed vowel identity were expected to affect F0 only slightly, since high vowels (such as in *Lilli*) have slightly higher F0 values than lower vowels (such as in *Lena*). The examples below illustrate one set of sentences, obtained by all combinations of the S and P phrase length manipulations.

short S, short P: *Lilli-Marlene ist eine berühmte Frau aus Suhl.*
(Lilli Marlene is a famous woman from Suhl.)
medium S, short P: *Lilli-Matthilda ist eine berühmte Frau aus Suhl.*
(Lilli Matthilda is a famous woman from Suhl.)
long S, short P: *Lilli-Matthilda Müller ist eine berühmte Frau aus Suhl.*
(Lilli Matthilda is a famous woman from Suhl.)
short S, long : *Lilli-Marlene ist eine berühmte Frau aus Baden-Württemberg.*
(Lilli Marlene is a famous woman from Baden-Württemberg.)
medium S, long P: *Lilli-Matthilda ist eine berühmte Frau aus Baden-Württemberg.*
(Lilli Matthilda is a famous woman from Baden-Württemberg.)
long S, long P: *Lilli-Matthilda Müller ist eine berühmte Frau aus Baden-Württemberg.*
(Lilli-Matthilda Müller is a famous woman from Baden-Württemberg.)

All target sentences were preceded by the same context sentence (*Auf dem Zettel steht geschrieben:* “On the paper it was written:”) ending with a colon in order to induce the speakers to produce a pause between the two sentences. Each sentence was presented 5 times in randomized order. In sum, the corpus consisted of 1320 observations (3 S constituents x 2 P phrases x 2 words x 5 repetitions x 20 speakers). The recordings were made in the anechoic room at ZAS and they lasted around 45 minutes for each speaker.

Based on the acoustic signal, the following were labeled: the onset and offset of the pause before the target sentence, the duration of the target sentence, the duration of the S constituent, the duration of the P phrase and the duration of the first word in the target sentence. The F0 peak was automatically detected as the F0 maximum within the first word of the sentence. Finally, we automatically measured the mean F0 value in the sentence-final syllable with word stress as an approximation to a final L%. We take this value to be a useful approximation to the baseline of the pitch range for each speaker (Ladd, 2008). This measurement was taken to rule out that effects of phrase length on the initial F0 peak might be merely due to an increase in the pitch register of the whole F0 contour. An example of F0 labeling is given in Figure 1.

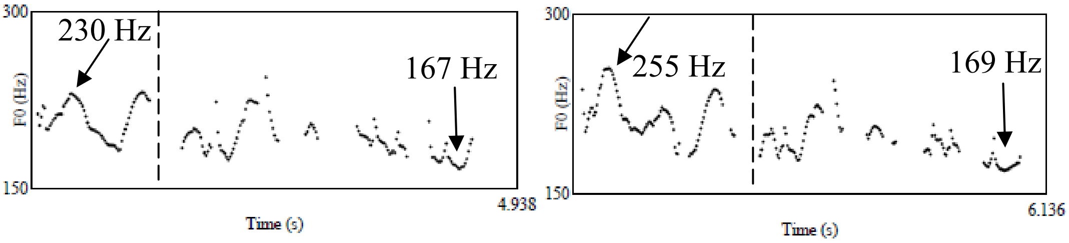


Figure 1. Labeling of the utterance-initial F0 peak and final F0 valley (indicated by the arrows) in a sentence with short S, long P (left) and with long S, long P (right). The dashed lines mark the end of the S constituent.

3. Results

Linear mixed models were run separately on pause duration and on the height of the utterance-initial F0 peak. For pause duration, the WM capacity (high/low), the length of the S constituent (short/medium/long), the length of the P phrase (short/long) and the Word type (Lilli/Lena) were included as fixed factors. For the initial F0 peak, Gender (male/female) was also included as the fifth fixed factor. Speakers were considered as random factors. Pairwise interactions between factors were also calculated, but they were factored out from the models when not significant. The cut-off point for significance was pMCMC < .01. The pMCMC refers to p-value calculated from a MONTE CARLO sampling by Markov chain, and it is commonly used in mixed models as an indicator of statistical significance (Fuchs et al., submitted).

3.1. Pause

Figure 2 shows pause duration data split across WM capacity, separately for the three S lengths. As it can be seen, differences across WM capacity are very small, with pause duration being slightly longer in speakers with low (0.23 s) than with high (0.18 s) WM capacity.

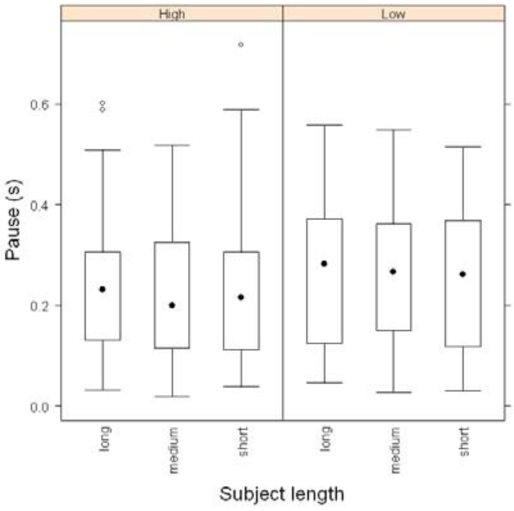
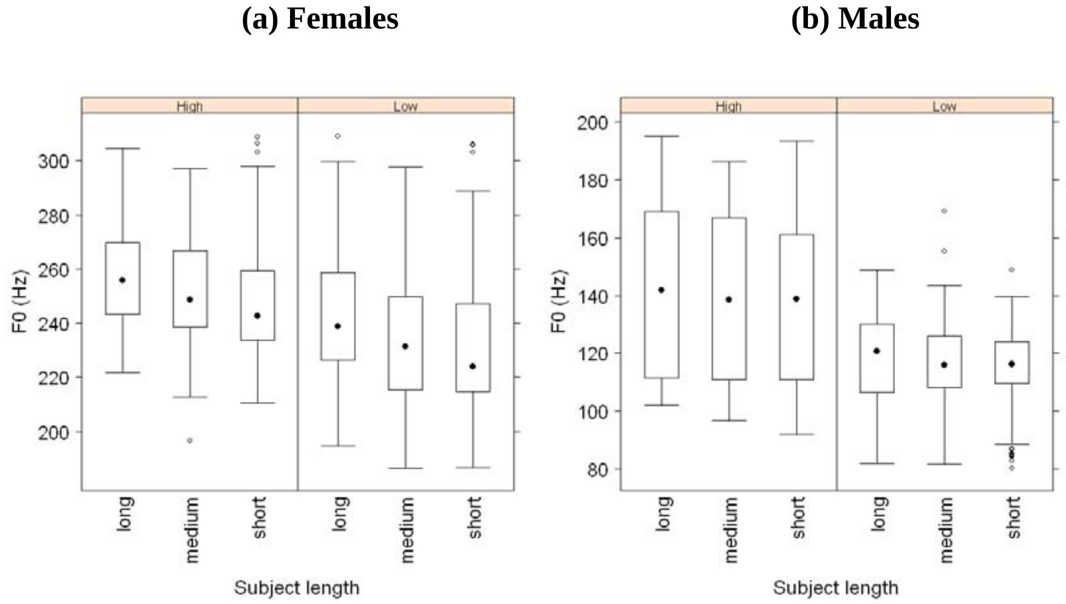


Figure 2. Boxplots for pause duration (s) against S length split by WM capacity. Data are collapsed across P length and word type.

Similar results were obtained for the two P lengths and their graphical illustration is thus omitted here for the sake of brevity. The statistical analysis showed no effects of length manipulations, neither for the S constituent nor for the P phrase. The effects of WM capacity, Word type as well as the interactions were also not significant.

3.2. F0

The height of the utterance-initial F0 peak was more sensitive to our manipulations. From Figure 3, it appears that speaker with high WM capacity show an overall higher F0 peak (mean in females: 253 Hz; males: 139 Hz) than speakers with low WM capacity (mean in females: 237 Hz; males: 116 Hz) across the S length manipulations. Moreover, female speakers with both high and low WM capacities tend to increase the initial F0 peak as the length of the S constituent increases. The greatest mean F0 difference (21 Hz) between long and short S constituents was produced by a female speaker with high WM capacity. As for the P length manipulation, longer P phrases led a few speakers to raise the initial F0 peak, but the mean difference between long and short P was only between 1-5 Hz.



(a) Females **(b) Males**

Figure 3. Boxplots for initial F0 peak height (Hz) against S length for females (left) and males (right) speakers. In each plot, results are split by WM capacity. Data are collapsed across P length and word type.

The statistical analysis performed on the utterance-initial F0 peaks showed a main effect of WM capacity [t = -1.6, pMCMC <.01]. Moreover, the contrasts among the S constituents were significant in female speakers for long vs. medium S [t = -4.168, pMCMC <.001]) and long vs. short S [t = -6.223, pMCMC <.001] but not for medium vs. short S [t = 2.086, pMCMC =.037]. Male speakers did not show any S effect, even though the contrast between long vs. short S was close to the cut-off point [t = -2.6, pMCMC =.013]. For both females and males, the contrast between long and short P was not significant. As expected, Gender [t = -7.8, pMCMC <.001] and Word type [t = 2.6, pMCMC <.001] turned out to be significant factors. Apart from the interaction between S length and Gender, all the other interactions were not significant. Finally, we found that, while the S manipulation had an effect on the utterance-initial F0 peak, it had no effects on the utterance-final F0 valley manipulations [pMCMC >.1].

4. Discussion

In this experiment we investigated the flexibility of the scope of planning from a phonetic point of view. We addressed the question whether speaker-specific variability varies with individual differences in WM capacities. Using a WM test and an acoustic task we showed that speaker-specific differences in intonation planning (as reflected in utterance-initial F0 peaks) might be related to WM differences, since speakers with high WM capacities start with higher F0 peaks than people with low WM capacities. This can be explained by the fact that, since speakers with high WM capacities plan larger chunks, they also have to start higher in order to apply F0 declination on larger portions of the utterance. One might wonder whether the variability in utterance-initial F0 height could be better explained in terms of an increase in the pitch register of the whole F0 contour due to paralinguistic factors, such as the emotional state of the speaker or conversational settings (Ladd, 2008). This hypothesis was ruled out, since in our experiment the value of the utterance-final F0 valley did not differ among the speakers and within the speaker, regardless of the experimental manipulations.

We also found that, when producing the utterance-initial F0 peak, speakers with both high and low WM capacity take into account the Subject length, but not the Prepositional phrase length manipulations. We interpret this result as indicating that the scope of intonation planning is local, with the F0 height being influenced by the length of the constituent in which it is realized. We also propose that such a constituent is not primarily syntactic in nature, but prosodic, since long Subject constituents in German (such as the ones employed in this acoustic task) are likely to be uttered as single prosodic constituents (see Fuchs et al., submitted for a discussion and some evidence based on auditory annotation).

Finally, we found no WM effects on pause duration. This was expected, since the analyses by Fuchs et al. (submitted) already showed that pause duration is not related to the length of the upcoming utterance, and that the cross-speaker variability is also very small. However, more research is needed to clarify why, contrary to the literature, we found no effects of sentence length on pause duration.

To sum up, the results of our study suggest that the scope of planning is speaker-dependent, thus supporting models of speech production which posit flexible units of planning. Moreover, our experiment reconciles contradictory findings in the literature on intonation planning by offering an explanation for the optional initial-pitch raising. Consequently, subject-dependent variability is not just the result of the mere physical actualization of fixed phonological categories. Explaining speaker-specific variability can lead to a better understanding of higher order cognitive behaviour and can shed some light on issues concerning phonological representation.

References

Conway A.R.A., M. Kane, M.F. Bunting, D. Zach Hambrick, O. Wilhelm & R.W. Engle Working memory span tasks: A methodological review and user’s guide. *Psychonomic Bullettin & Review*, 12 (5), 769-786, 2005.

Cooper, W. E. & Paccia-Cooper, J. *Syntax and speech.* Harvard U. Press, 1980.

Ferreira, F. & Swets, B. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *J. Mem. and Lang.* 46, 57–84, 2002.

Fuchs, S. , Petrone, C., Krivokapić, J. & Hoole, P.. Is utterance planning reflected in acoustic and respiratory parameters? (submitted)

Goldman-Eisler, F. *Psycholinguistics. Experiments in spontaneous speech.* London and New York: Academic Press, 1968.

Krivokapić, J. Prosodic planning: Effects of phrasal length and complexity on pause duration. *JPhon* 35, 162–179, 2007.

Ladd, D.R. *Intonational phonology.* Cambridge: CUP, 2008.

Levelt, W. J. M. *Speaking. From intention to articulation.* Cambridge: MIT Pr, 1989.

Prieto, P., D'Imperio, M., Elordieta, G., Frota, S. & Vigário, M. Evidence for soft preplanning in tonal production: Initial scaling in Romance. *Proc. of Speech Prosody*. Dresden: TUDpress, 803–806, 2006.

Swets, B., Desmet, T., Hambrick, D. Z. & Ferreira, F. The role of working memory in syntactic ambiguity resolution: A psychometric approach. *J. Exp. Ps.: General* 136, 64–81, 2007.

Wagner, V.; Jescheniak, J. D. & Schriefers, H. On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *J. Exp. Psychology: Learning, Memory, and Cognition,* 36, 423-440, 2010.

Wheeldon, L. & Lahiri, A. Prosodic units in speech production. *J. Mem. and Lang.* 37, 356–381, 1997.

Zvonik, E. & Cummins, F. The effect of surrounding phrase lengths on pause duration. *Proc. .Eurospeech*. Geneva, Switzerland, 777–780, 2003.