

## **Memory Demands in Linguistic Compensation**

Srdan Medimorec

Chandeep Mander

Evan F. Risko

University of Waterloo

Corresponding author information:

Srdan Medimorec

Department of Psychology, University of Waterloo, 200 University Avenue West, Waterloo ON,

Canada, N2L 3G1

[smedimorec@uwaterloo.ca](mailto:smedimorec@uwaterloo.ca)

+1 519-781-3691

Word count: 4,549

## Abstract

Individuals often modify speech characteristics to accommodate their listeners. In the present study we investigate how speakers modify their speech in a dictation task and what this says about their beliefs with respect to the listener's information processing limitations. To do so, we asked participants to either read a set of numbers aloud, or dictate numbers so that another person could write them down. Our results suggest that speech modification in this task was not related to the speakers' differences in working memory capacity, and could represent speakers' attempt to minimize working memory demands of the listener. This account is similar to minimal memory strategies, whereby individuals often try to minimize memory demands in cognitive tasks.

Keywords: language; linguistic compensation; speech; working memory

### **Memory Demands in Linguistic Compensation**

Talkers modulate speech characteristics as a function of context (Arnold, Kahn, & Pancani, 2012; Ashburn & Gordon, 1981; Medimorec, Pavlik, Olney, Graesser, & Risko, 2015; Pelegrín-García, Smits, Brunskog, & Jeong, 2011; Rantala, Hakala, Holmqvist, & Sala, 2015; Walker, Risko, & Kingstone, 2014). For example, speakers use more fillers (*um* and *uh*) when addressing a human than a computer (Walker et al., 2014), talk more loudly in noisy conditions (Rantala et al., 2015), and decrease speech rate when speaking to a distant person (Pelegrín-García et al., 2011). Such speech modulations are usually interpreted to indicate a presence of communicative intent (Cooke, King, Garnier, & Aubanel, 2014; Garnier, Heinrich, & Dubois, 2010; Medimorec et al., 2015). In the current study we expand on the investigation of context induced compensatory speech modification by comparing speech modulations across two contexts, namely dictation and reading aloud. Dictation represents a straightforward task that should elicit an online speech modification, thus providing us an opportunity to investigate, under controlled conditions, the ways in which individuals adjust their speech, and the potential mechanisms underlying these adjustments.

#### *Compensation in Speech*

According to previous research, speakers' engagement in compensating linguistic behavior reflects communicative strategies aimed at improving listeners' comprehension (Cooke et al., 2014; Medimorec et al., 2015). For example, Medimorec et al. (2015) demonstrated that lecturers compensate for the use of abstract words by using more cohesive and syntactically simpler discourse. Another form of speech modulation that is often adopted by talkers is pause insertion (Bradlow, Kraus, & Hayes, 2003; Gustafson-Capkova & Megyesi, 2001; Picheny, Durlach, & Braida, 1986; Smiljanić & Bradlow, 2005; Swerts & Geluykens, 1994). For example,

Bradlow et al. (2003) asked participants to read sentences in their conversational speaking style (i.e., like reading to a listener who is highly familiar with their voice), or in a clear speaking style (i.e., like reading to a listener with hearing difficulties, or a listener of a different language background). The authors reported that speakers increased pause rates in the clear speaking condition compared to the conversational speaking style condition, presumably reflecting speakers' response to their audience (i.e., an audience design; Bell, 1984). Thus in this case speakers engaged in a deliberate manipulation of discourse in order to accommodate their audiences' perceived (or implied) processing limitations. On the other hand, in some instances speech modification might represent a byproduct of demands on production, rather than a product of controlled processing. For example, the use of abstract concepts in speech might increase the cognitive load on the speaker, forcing the adoption of simpler syntax (e.g., Medimorec et al., 2015). Regardless of the underlying mechanism, it is worth noting that there is evidence that speech modulations can indeed benefit listeners (Bradlow, et al., 2003; Picheny, et al., 1986). For example, Bradlow et al. (2003) demonstrated that an increased number of pauses in speech improved comprehension in children with learning disabilities. Increased pause rates were demonstrated to have similar benefits when speech is directed to hearing-impaired adults (Picheny et al., 1986) or in adverse conditions such as noisy environments (Garnier et al., 2010).

### **Present Investigation**

As noted above, much previous research has investigated how individuals' speech modulation varies as a function of communicative intent (e.g., Cooke et al., 2014; Medimorec et al., 2015). We expand on these findings by comparing speech characteristics across two conditions: dictating and reading aloud. While previous research has studied dictation in the

context of speech recognition technologies (e.g., De La Paz, 1999; MacArthur, & Cavalier, 2004), we explore the utility of dictation in the context of linguistic compensation.

In the current study we asked participants to either read aloud sets of random numbers, or dictate the numbers to another person. While we might expect participants to do what is easiest for them in the read condition (e.g., pause less frequently, since there is no communicative purpose), any potential speech modulation in the dictate condition should reflect an audience design. With the exception of their audiences' task (i.e., write the numbers down), participants were not aware of any of the addressees' communicative needs (e.g., writing speed), since they were not able to see their audiences' writing nor did they receive any verbal feedback. In an attempt to use a relatively decontextualized task we use sets of numbers rather than actual sentences (i.e., context might influence the prosody of utterances, for example words predictable in context are often acoustically reduced; Arnold et al., 2012).

Thus in the current study we investigate whether individuals change their speech characteristics in the dictation task to accommodate the listener by, for example, making more pauses (and conversely uttering shorter strings of numbers). Assuming they do, we were interested in the nature of this accommodation. In a dictation task being on the listener side (i.e., the writing side) is similar to performing a kind of working memory task (i.e., keeping numbers in mind while writing them down). Thus, the listener's working memory (WM) capacity represents a critical constraint on successful dictation. For example, dictating 7 digits followed by a pause would likely overburden most people's WM capacity, given that individuals WM capacity is limited to about 3-5 items (Cowan, 2010; Unsworth & Engle, 2006). As such the number of items uttered between pauses provides some indication of the speaker's expectations

about the WM capacity of the listener. Understanding how the speaker sets this expectation would provide important insight into context induced compensatory speech.

Here we contrast a number of potential hypotheses. If individuals used their own WM capacity as a proxy for other's WM capacity, then we would expect individuals with higher WM spans to utter more digits before pausing compared to individuals with lower WM spans. On the other hand, rather than using themselves as a model, individuals might instead have a mental model of other's WM limitations (e.g., acquired through experience interacting with others), in which case there would be no relation between an individual's WM capacity and the number of digits they utter before pausing. In the case that individuals have such a mental model the number of digits between pauses would reflect individual's accommodation for listener's perceived information processing limitations. Here we consider two plausible forms that such accommodation could take. First, individuals could use a kind of "average" WM capacity (based on their everyday experience, for example when remembering a new phone number, or calculating something "in the head") and, assuming that working memory is limited to about 3-5 items (Cowan, 2010), utter 3-5 items before pausing. Alternatively, individuals could take a kind of "minimal memory" or "lowest common denominator" approach by uttering a number smaller than what would be expected based on the average WM capacity (e.g., by uttering 1-2 items before pausing). This latter account is similar to minimal memory strategies proposed by Ballard, Hayhoe, Pook, and Rao (1997), whereby individuals often try to minimize memory demands in cognitive tasks. For example, Risko and Dunn (2015) reported that in a short-term memory task when given the option of writing down to-be-remembered information participants did so almost 50% of the time when only two letters had to be remembered, which is well below most individuals' capacity. Thus many individuals appeared to be minimizing WM resources. While

conceptually similar, here the locus of WM savings shifts to the listener. For example, uttering digits one-by-one would suggest that speakers tried to maximally "offload" their audience's WM demands (Risko & Gilbert, 2016). In addition to the size of digit groupings, another form of compensatory linguistic behavior in this task would be to increase pause duration with increases in the size of digit groups. For example, if speakers are trying to accommodate their audience, then larger groups of digits should be followed by longer pauses, giving the listener enough time to perform the task (i.e., write down the numbers being held in mind).

Thus in the current study we test the prediction that individuals should modify their speech by pausing when dictating numbers, compared to only reading the numbers aloud. We investigate whether such changes in speech are related to working memory capacity, or whether speakers try to maximally offload their listener's working memory demands.

## **Method**

*Participants.* Participants were 82 university students. One participant refused to be recorded, and 2 participants did not follow the instructions, so the final sample included 79 participants. All participants were compensated with course credit.

*Design.* A within-subjects design with two conditions (read vs. dictate) was used. Each participant was asked to read and dictate strings of numbers (15 strings of 24 numbers per condition). The order of conditions was counterbalanced (i.e., the order of the conditions changed for consecutive participants across the experiment). The dependent variables were digit groupings (i.e., the average number of digits uttered between pauses) and pause durations (in ms) across the condition.

*Stimuli and Apparatus.* Each number string was printed on an individual piece of paper. There were 15 sets of randomly generated number strings, each containing 24 numbers, ranging

from 1-9. Numbers were printed in Verdana 12pt font, and consecutive numbers were separated by a comma and a blank space. The sets were numbered and placed in a small container.

Participants' responses were recorded by using a digital voice recorder (Olympus VN-702PC).

WM capacity was measured by using an automated version of the operation span task (OSPAN; Unsworth, Heitz, Schrock, & Engle, 2005). In this task, participants are asked to solve a math problem, followed by a presentation of a letter that needs to be remembered. Set sizes (math problem-letter pairs) range from 2-5 and are randomly presented (each set size is presented for 3 trials). At the end of each trial, participants are asked to report the presented letters. There is an 85% accuracy criterion for math operations (participants below this criterion are not scored). The WM capacity measure reported here represents the average number of items participants can hold in WM. This number was calculated by dividing the absolute OSPAN score (representing the sum of correctly recalled letter sets in order of presentation) by 12 (3 trials x 4 set sizes).

*Procedure.* Participants were seated with their back to a research assistant. In the reading condition participants were told that there are 15 sets of random numbers in the container. Participants were instructed to draw the sets out one by one, state the set number, and then to read the numbers aloud. In the dictation condition participants were told that there are 15 sets of random numbers in the container, they were instructed to draw the sets out one by one, state the set number, and then dictate the numbers so that the research assistant could write them down. Participants were given no further instruction. The research assistant was instructed not to give any feedback regarding dictating (or reading) speed. Participants' responses were audio recorded. After completing the task, participants completed the automated OSPAN test. Our version of the test was administered using E-Prime 2.0 software. The test took an average of about 13 minutes to administer.

*Pauses in Speech.* For the purposes of the current study pause threshold in speech (onset of the current number-offset of the previous number) was set to 300 ms (e.g., ten Bosch, Oostdijk, & Boves, 2005; Wang et al., 2012). The choice of this pause threshold was also based on a pilot study. We asked three independent coders to detect the number of pauses using a sample of 30 participants in reading ( $N = 15$ ) and dictate conditions (each participant read or dictated a single string of 12 numbers). The average number of detected pauses across the coders correlated highly with the number of pauses of 300 ms or more ( $r(28) = .86, p < .001$ ).

Pauses were determined automatically, using Audacity software (version 2.1.0) and were checked manually for accuracy.

## Results

In the current study digit grouping sizes represent the average number of digits uttered between pauses (of 300 ms or more). First we calculated the number of digit groups per string (number of pauses + 1). A digit grouping size was determined by dividing 24 (i.e., 24 numbers per string) by the number of digit groups per string. Since digit grouping sizes (and pause durations) were not normally distributed in any condition (read or dictate; the Shapiro-Wilk statistic, digit grouping  $ps < .001$ , pause duration  $ps < .001$ ), we performed a within subject non-parametric test (i.e., Wilcoxon Signed-Rank) to assess the effect of condition.<sup>1</sup> Reported effect sizes ( $r$ ) are calculated by dividing the Z value by the square root of N (the number of observations over the two conditions; Pallant, 2007). These effect sizes represent Cohen (1998) d-like estimates. Condition order did not have an effect nor interacted with any variable of interest (i.e., digit groupings and pause durations). Reported correlations are Spearman's rho ( $r_s$ ).

---

<sup>1</sup> Results were qualitatively similar when we used repeated measure ANOVAs to analyze the log10 transformed data.

*Digit Groupings and Pause Durations Across the Condition*

Condition had a significant effect on the average number of digits uttered without pausing, such that there were less digits uttered in the dictate condition compared to the read condition,  $Z = -6.42$ ,  $p < .001$ ,  $r = .51$ . Moreover, pauses between digit groups were longer in the dictate condition compared to the read condition,  $Z = -6.00$ ,  $p < .001$ ,  $r = .48$  (for means and standard deviations see Table 1). Thus, the need to communicate with a partner clearly lead to a marked change in language behavior.

Table 1

*The mean number of digits uttered without pausing, pause durations and standard deviations across two conditions*

Condition	Digit Groupings		Pause Durations (ms)	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Read	4.89	4.20	554	187
Dictate	2.46	2.38	773	361

*Digit Grouping Consistency*

Next we investigated the consistency of individuals' digit grouping patterns. Here we performed bivariate correlations between digit grouping sizes, and between pause durations in even and odd sets (or number strings) for each participant in the dictate condition (i.e., we split each participant's trials into two sets, computed the average digit grouping and pause duration in each of the sets, and correlated those values). The correlation between digit grouping sizes in even and odd sets was high,  $r_s(77) = .97$ ,  $p < .001$ . The same was true for the correlation between

pause durations in even and odd sets,  $r_s(77) = .96, (.95) p < .001$ . Thus, participants displayed consistent digit grouping strategies and pausing throughout the dictation task.

#### *Digit Groupings and Pause Durations Within an Individual*

Next we examined the relation between average digit grouping sizes and pause lengths within an individual. Pause lengths represent the mean pause duration per number string (including only pauses of 300 ms and above). We calculated correlations between grouping sizes and pause durations for each participant across 15 number sets. Thus, we split the file by participant and calculated a single correlation coefficient for each participant relating their grouping sizes and pause durations. When we compared these coefficients to zero, the analysis demonstrated a positive relation between digit grouping sizes and pause durations,  $t(78) = 4.13, p < .001$  ( $M = .17, SD = .36$ ). Thus, larger grouping sizes were related to longer pauses within an individual. When the same analysis was performed in the reading condition, there was no statistically significant relation between digit grouping sizes and pause durations,  $t(78) = .89, p = .379$  ( $M = .04, SD = .37$ ).

Finally, correlations between grouping sizes and pause durations were significantly higher in the dictate condition compared to the read condition,  $t(78) = 2.54, SEM = .05, p = .013$ , strong evidence of compensating behavior in the dictate condition.

#### *Digit Grouping Behavior*

Next we explored the way in which individuals grouped digits in the dictate condition. As is clear from Table 2, the majority of participants (86.1%) in this condition used digit groups ranging from 1 to 3 digits. Also, the most popular grouping was 1 digit (~47%). As is clear from Table 2, a subset of participants (i.e., those with digit groupings  $\geq 6$ ) do not appear to have engaged in any compensatory pausing behavior in the dictate condition. Instead the groupings

likely reflect taking a breath given the length of the string. As such, in the following analyzes we report both a full sample analysis and an analysis restricted to participants with digit groupings of 4 or less in the dictate condition (92.4% of the sample).

Table 2

*Distribution of digit grouping sizes (i.e., digits uttered without pausing), working memory capacity, and mean pause durations*

% of Participants	Grouping Size	Working Memory Capacity	Pause Duration (ms)
		<i>Mean (SD)</i>	<i>Mean (SD)</i>
46.84	1	3.07 (1.22)	872 (367)
21.52	2	3.11 (1.04)	729 (392)
17.72	3	3.13 (1.21)	777 (276)
6.33	4	2.07 (1.62)	542 (216)
7.59	>=6	2.94 (.88)	478 (64)

Note: Grouping sizes are rounded to the nearest whole number

#### *Digit Groupings, Pause Durations, and Working Memory Capacity*

The average WM capacity (i.e., the number of items individuals could hold while performing the math task) was 3.03 ( $SD = 1.21$ ,  $Range = 4.60$ ). When we restricted the sample to participants with digit groupings of 4 or less in the dictate condition, the average WM capacity was 3.78 ( $SD = 1.51$ ,  $Range = 5.75$ ). To determine the relation between WM capacity and digit grouping we computed the bivariate correlation between the WM capacity measure and average digit groupings in the dictate condition. The full sample analysis is based on 78 participants (one

participant failed to meet the 85% math test accuracy). There was a negative, although not statistically significant correlation between the two measures,  $r_s(76) = -.11, p = .354$ . When we included only participants with digit groupings of 4 or less in the dictate condition, the correlation remained non-significant and in the same direction,  $r_s(70) = -.10, p = .445$ . As is clear from Table 2, individuals' WM capacity and grouping sizes were not systematically related. In addition, WM capacity was not related to pause durations,  $r_s(76) = .12, p = .314$ . The same was true for the participants with digit groupings of 4 or less,  $r_s(70) = -.10, p = .412$ . Similarly, the within subject correlation between digit grouping size and pause duration was also not related to WM capacity,  $r_s(76) = .02, p = .868$ . The same was true for the participants with digit groupings of 4 or less,  $r_s(70) = .02, p = .891$ .

### Discussion

The current study contrasted different hypotheses about compensatory speech modification in a dictation task. Our straightforward prediction was that individuals would modify their speech by inserting more pauses when dictating numbers to another person, compared to only reading them aloud. In addition, we investigated whether individuals used their own WM capacity as a proxy for other's WM capacity, or accommodated their listener by uttering what might be considered an "average" WM capacity before pausing (3-5 items, reflecting the average WM constraint; Cowan, 2010), or used a kind of minimal memory approach by uttering 1-2 items before pausing. Our analysis demonstrated that individuals used smaller number groupings, and longer pauses in the dictate condition compared to the read condition. In addition, individuals displayed a consistent digit grouping and pausing strategy throughout the dictation task. Moreover, pause durations were related to digit grouping sizes

within an individual such that that larger digit groupings were followed by longer pauses, and this relation was significantly higher in the dictate condition compared to the read condition.

Consistent with the minimal memory approach prediction, the majority of participants (~68%) uttered numbers one-by-one or in groups of two, well below (a) individuals' WM capacity (here ~ 3.8 items), and (b) what might be considered an average WM capacity (i.e., 3-5 items; Cowan, 2010). Furthermore, compensatory behavior (i.e., number grouping and pause durations) was not related to participants' WM capacity. Thus individuals engaged in a form of linguistic compensation in our tasks and this compensation appeared to take a kind of "minimal memory" form.

#### *Potential Mechanisms Underlying Speech Modification in the Current Study*

The instruction to dictate numbers to another person (versus read them aloud) markedly changed the way that the numbers were uttered. For example, the average digit grouping size decreased from 4.9 to 2.5. Moreover, the results of the current study demonstrated that individuals in the number dictation task tended to adopt what could be considered a minimal memory approach, provided they most commonly uttered digits one-by-one or in groups of two. In addition, individuals were highly consistent in the use of similar grouping patterns throughout the dictation task, suggesting that a decision about the "optimal" grouping size strategy is robust and reached early in the task. One potential explanation of the observed speech patterns is that individuals attempt to maximize listener accuracy and/or decrease listener effort. Such a strategy might be driven by a lack of information about their addressee. For example, since participants were not able to see how their addressee was performing the task nor get any feedback, there was no reliable information that could be used to adjust characteristics of speech (e.g., digit groupings). In such a case adopting a minimal memory approach (from the onset of the task) by

uttering number groupings smaller than what would be expected based on an approximation of people's average WM capacity seems prudent. This approach would accommodate the majority of potential addressees (indeed in our sample the average WM capacity was about 4 items).

While previous research has reported that individuals adopt the minimal memory approach when it can potentially benefit their performance (e.g., writing down to-be-remembered items, instead of keeping them in the head; Risko & Dunn, 2015), the finding reported here demonstrates that individuals may use a similar strategy in an attempt to minimize another persons' cognitive demands.

Another potential indicator of listener oriented compensational speech modification reported here is the use of longer pauses after uttering larger number grouping sizes within an individual. A straightforward interpretation of this result is that speakers attempted to accommodate listeners by giving them more time after larger number groupings. In this case, systematic variation in pause durations arguably demonstrates individuals' awareness of their listener's task (i.e., writing numbers down) throughout the experiment.

#### *Future Directions*

While the results of the current study clearly support the notion that individuals modify speech during dictation, it is also worth noting that about 8% of the participants did not seem to do so (i.e., the participants with digit groupings of 6 or more, see Table 2). Whether such a behavior may be predictable on the basis of personality traits (e.g., conscientiousness) remains an interesting question to further pursue. Moreover, in the current study participants were unaware of the listener's needs. The interesting question thus remains whether being aware of the listener's demands would result in a different behavior. Further research introducing such a

"listener-aware" condition could provide a better understanding of the role that awareness (or unawareness) to the listener's needs plays in adjusting one's speech.

*Conclusion*

The current study systematically investigated context induced compensatory speech and revealed a number of interesting results. Our results demonstrated that speakers, regardless of their WM capacity, set their expectations about the WM capacity of the listener well below the average WM capacity in a situation where such information is ambiguous. Thus, decisions about how to structure speech appears aimed at maximally offloading listeners' resources.

## References

- Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*, *19*, 505-512.
- Ashburn, G., & Gordon, A. (1981). Features of a simplified register in speech to elderly conversationalists. *International Journal of Psycholinguistics*, *8*, 7-31.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*, 723-742.
- Bell, A. (1984). Language style as audience design. *Language in Society*, *13*, 145-204.
- Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, *46*, 80-97.
- Campbell, T. F., & Dollaghan, C. A. (1995). Speaking rate, articulatory speed, and linguistic processing in children and adolescents with severe traumatic brain injury. *Journal of Speech, Language, and Hearing Research*, *38*, 864-875.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates
- Cooke, M., King, S., Garnier, M., & Aubanel, V. (2014). The listening talker: A review of human and algorithmic context-induced modifications of speech. *Computer Speech & Language*, *28*, 543-571.
- Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why?. *Current Directions in Psychological Science*, *19*, 51-57.

- De La Paz, S. (1999). Composing via dictation and speech recognition systems: Compensatory technology for students with learning disabilities. *Learning Disability Quarterly*, 22, 173-182.
- Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research*, 53, 588-608.
- Gustafson-Capkova, S., & Megyesi, B. (2001). A comparative study of pauses in dialogues and read speech. In *INTERSPEECH* (pp. 931-934).
- MacArthur, C. A., & Cavalier, A. R. (2004). Dictation and speech recognition technology as test accommodations. *Exceptional Children*, 71, 43-58.
- Medimorec, S., Pavlik Jr, P. I., Olney, A., Graesser, A. C., & Risko, E. F. (2015). The language of instruction: Compensating for challenge in lectures. *Journal of Educational Psychology*, 107, 971-990.
- Pallant, J. F. (2007). *SPSS survival manual: A step-by-step guide to data analysis with SPSS*. New York, NY: McGrath Hill.
- Pelegrín-García, D., Smits, B., Brunskog, J., & Jeong, C. H. (2011). Vocal effort with changing talker-to-listener distance in different acoustic environments. *The Journal of the Acoustical Society of America*, 129, 1981-1990.
- Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, 29, 434-446.
- Rantala, L. M., Hakala, S., Holmqvist, S., & Sala, E. (2015). Classroom Noise and Teachers' Voice Production. *Journal of Speech, Language, and Hearing Research*, 58, 1397-1406.

- Risko, E. F., & Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and Cognition*, *36*, 61-74.
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*, 676-688.
- Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, *118*, 1677-1688.
- Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and speech*, *37*, 21-43.
- ten Bosch, L., Oostdijk, N., & Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Communication*, *47*, 80-86.
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104-132.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498-505.
- Walker, E. J., Risko, E. F., & Kingstone, A. (2014). Fillers as signals: Evidence from a question–answering paradigm. *Discourse Processes*, *51*, 264-286.
- Wang, Y. T., Nip, I. S., Green, J. R., Kent, R. D., Kent, J. F., & Ullman, C. (2012). Accuracy of perceptual and acoustic methods for the detection of inspiratory loci in spontaneous speech. *Behavior Research Methods*, *44*, 1121-1128.