THE ABSENCE OF PAUSES IN SPOKEN NARRATIVES AND MEMORY RECALL

Jan Volín and Pavel Šturm

Institute of Phonetics, Charles University, Prague, Czech Republic
jan.volin@ff.cuni.cz, pavel.sturm@ff.cuni.cz

ABSTRACT

An experiment with a memory recall test based on continuous spoken texts was carried out. The primary factor to examine was the absence of pauses. A sample of 24 subjects listened to eight narratives with the task to subsequently fill in the missing words into gaps in a printed version of each of the texts. With pauses defined as cessation of lexical item articulation, the present study tests the hypothesis that spoken texts with the pauses removed will lead to a worse recall of the words contained in them.

The results indicate that pauses have an impact on the effectiveness of language communication, as the lexical content of spoken texts without pauses was more difficult to retrieve relative to texts with the pauses intact. This seems to suggest that apart from speech production requirements, pauses meet beneficially some of the perceptual constraints. Nevertheless, the resulting effect was neither straightforward nor monotonous across texts. Benchmark data for further experimenting are provided.

Keywords: pause, memory, recall, prosodic structure, temporal organization.

1. INTRODUCTION

One of the central questions in linguistics is – or if it is not, perhaps it should be – that of effectiveness of communication. The costs of achieved correspondence between the speaker's intended meaning and the addressee's recovered meaning seem to be an obvious indicator of the effectiveness. However, there is currently no procedure that would allow for measureing them directly. Proxies that are often used instead include the complexity of cerebral processing as reflected in reaction times while fulfilling various behavioural tasks, scaled judgements of the speaker's personality features (e.g., agreeableness, competence) or, finally, various measures of memory performance. This latter one will be focused on in the present study.

Memory recall has been tested prevalently on strings of individual items: numerals, unrelated nouns, non-words, etc. (e.g., [1], [2], [3], [4], [5], [6]). Yet, it is undisputed that natural language communication uses utterances as primary structural units, even if occasionally very short ones. Utterances can be analysed as fields of relations among words [7] and relational links outside the utterance have also been often considered. Lists of isolated items lack this essential feature of everyday language use. The authors in [6] actually state that "…memory for lists of words likely differs from memory for conversational speech in important ways". Therefore, we decided to test the listeners' memory for the contents of spoken texts using continuous strings of utterances in narratives.

It is widely acknowledged that the impact of a message is not attributable solely to the choice of words or syntactic links among them. The prosodic form is seen as an important factor in speech comprehension modelling (e.g., [8], [9], [10], [11]). Interestingly, in [12] an effect of such a crude measure as mean F0 (a correlate of mean pitch termed *key* or *register*) on memory was established with clear, albeit complex results.

Our study focuses on a prosodic event that is also quite salient: the pause. Superficially, pauses are viewed as a response to the plain need to breath, but they actually reflect certain cognitive constrains. First, speakers need pauses during neuro-linguistic planning of their speech unit [13]. Second, and more pertinent to our study, listeners may need pauses to process the contents of utterances and to store information in their memory (see, e.g., [14]).

In [5], for instance, the impact of pause presence on number recollection was investigated. The authors extended their previous research that suggested a beneficial effect of pauses on remembering the contents of numeral strings. Apart from confirming their previous findings, the new experiment showed that longer pauses were better than the shorter ones.

The chief objective of our study is to determine how absence (rather than presence) of pauses affects memory for lexical contents of narratives. Since presence and absence are complementary modes, this wording might deserve explanation. In the previously mentioned study [5], the researchers added pauses into their speech material. We, on the other hand, carefully deleted naturally occurring pauses from our narratives (see Section 2).

Although the null hypothesis would stipulate no difference between our two conditions (*Pause* vs. *NoPause*), we expect words from the spoken texts with the pausing left untouched to be better recalled.

2. METHOD

2.1. Material

Eight extracts from narratives were used for the experiment. They were all obtained from high-quality audiobooks produced in professional studios with experienced actors (4 women + 4 men). The general style could be characterized as fluent speech read out from a well-edited written text. The extracts were selected so that no markedly tense atmosphere nor overly dynamic development of the plot was present. Therefore, we assume that topics of the narratives were of no crucial importance.

As the narrative recordings were made in studios, there were only logical pauses in line with syntactico-semantic structure (guaranteed by the studio director), that is, no abrupt or disturbing dysfluencies.

Mean duration of the extracts was 47.2 seconds with std. dev. of 1.1 seconds. Further descriptors of the sample are displayed in Table 1.

Speaker	DurP+	DurP–	n Wds	n Pss
F1	45.2	37.2	111	13
F2	46.3	37.1	119	14
F3	48.0	41.2	102	19
F4	46.4	34.3	91	15
M1	48.9	35.9	106	17
M2	46.7	38.1	116	15
M3	47.2	36.9	112	23
M4	48.3	37.3	84	18
Mean	**47.2**	**37.2**	**105.1**	**16.8**
Std.dev.	**1.1**	**1.8**	**11.5**	**3.0**

Table 1: Descriptive facts on the narrative extracts used in the test. DurP+ = Duration of the complete extract in seconds; DurP– = Duration with pauses removed (s); n Wds = number of words; n Pss = number of pauses.

Table 1 reveals that although the spoken text durations (whether with or without pauses) were quite balanced across the sample, the number of pauses ranged from 13 to 23. Due to the repeated-measure design of the study, other differences (tempo, liveliness, etc.) were not considered crucial.

Pauses were carefully removed in Adobe Audition CS6 sound editor with special attention to the resulting smooth transitions. In other words, sharp cut-offs or other sound artefacts were not allowed. Note that all other prosodic signals such as phrase-final lengthening (deceleration) were left untouched.

Two counterbalanced versions of a perception test were created in which half of the extracts had naturally occurring pauses and the other half had pauses removed. Listeners were assigned to the two groups randomly.

2.2. Participants

Volunteers from various philological programmes at the Faculty of Arts, Charles University were recruited to participate. There were 18 female and 6 male students (n = 24). Their mean age was 20.4 years and they reported no hearing nor dyslexic problems. In all the cases, the Czech language was their mother tongue. The volunteers expressed their wish to participate in a 'strictly anonymous listening test investigating how people remember individual words in story-telling'. They were blind to the experimental condition and a post-test interview revealed that they found the test interesting but difficult.

2.3. Testing procedure

The participants were invited in groups of 3 or 4 into a quiet room with a quality sound equipment. They were informed that extracts from spoken narratives would be played for them to listen to. After that, they would receive a printed transcript of each narrative with 10 words in each extract missing. Their task was to fill in the missing words into the gaps in the written text. Two spoken texts were presented at a time. Then the printed sheet for the first text would be handed out and, after completion, the second text would be done. The time limit for one text completion was 2 minutes.

At the end of each allotted time, the respondents were instructed to indicate whether they recognized the text by ticking YES or NO at the bottom of their answer sheet. The texts were generally unknown, only exceptionally the participants knew from which book the extracts came.

Before a new pair of extracts was played, the experimenter asked the participants about their feelings and encouraged them for the next stage. The whole testing session took 25 minutes (30 minutes with the arrival, greetings and the instruction).

2.4. Predictability without the sound

Before the actual perception test, the written texts were piloted with 8 different volunteers (of the same social characteristics as the perception test participants) to examine the predictability of the target lexical items. In other words, we tested which of the gaps could be filled in without listening to the spoken narrative.

The results of this pre-test showed that most of the words were utterly unpredictable (63/80 = 78.8%). Eight out of the remaining 17 words were correctly 'guessed' by just one of the subjects. Only three words (3/80 = 3.8%) manifested predictability over 50%, that is, more than half of the subjects filled them in correctly. These results were later taken into account in the course of the analyses (see below).

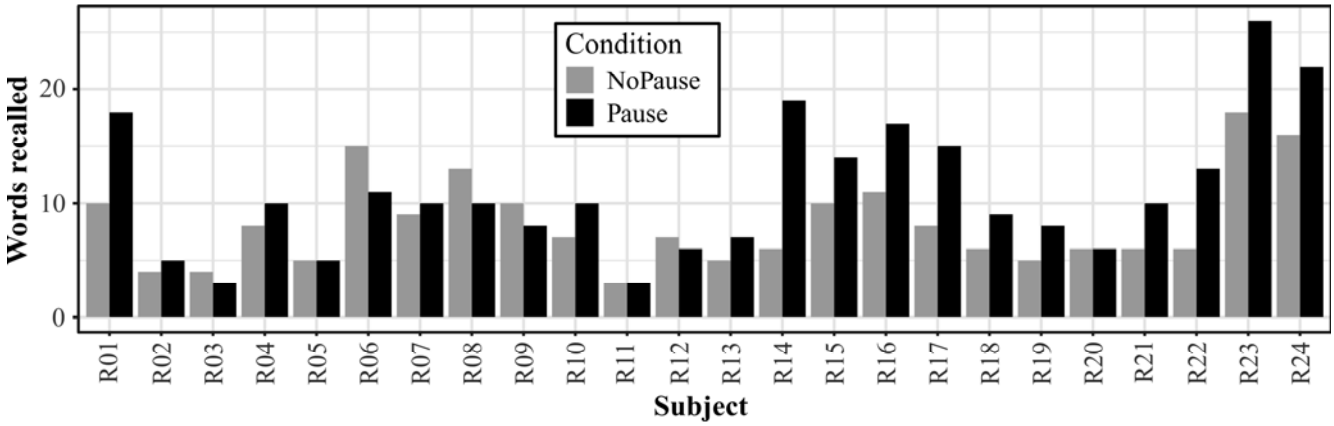

Figure 1: Counts of correctly recalled lexical items by 24 respondents (R01–R24) under two experimental conditions.

3. RESULTS

Due to the intended difficulty of the task, only 463 out of 1920 items (24 subjects × 80 words) were recalled correctly. In terms of percentage, it is 24.1% success rate. When split by CONDITION, the recall rates were 20.6% under the *NoPause* condition and 27.6% under the *Pause* condition (198 against 265 correctly recalled items, respectively). There were 10 words to be recalled per text. On average, 2.47 words were remembered correctly under the *NoPause* condition, whereas under the *Pause* condition it was 3.31 words.

Out of 40 missing words under each condition, an average respondent remembered only 8.25 under the *NoPause* condition, compared to 11.04 under the *Pause* condition. However, Figure 1 shows that individual respondents contributed to the results in a disparate manner. Whereas R23 and R24 are clearly champions of the sample, R03 and R11 had extremely poor recall. Be that as it may, two thirds of the respondents (16 out of 24) performed in the expected direction, that is, they achieved better results under the *Pause* condition. In contrast, there were also 5 subjects who performed in the opposite direction and 3 who had equal performance under both conditions.

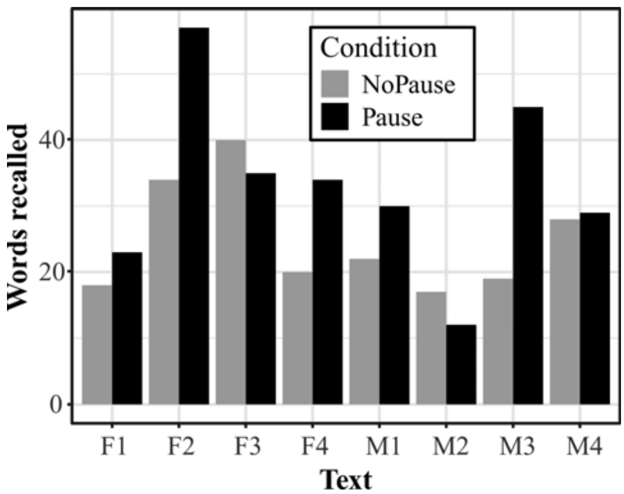

Figure 2: Counts of correctly recalled lexical items in 8 spoken texts under two experimental conditions.

The results can be further broken down according to the individual experimental texts. Figure 2 shows that the effect size varies across the eight texts and two of the texts (F3 and M2) even produced results in an unexpected direction.

To evaluate the difference between *Pause* and *NoPause* conditions statistically, we constructed a generalized linear mixed model with the Poisson distribution. The situation in Figures 1 and 2 suggests that the effect should vary freely across subjects and texts in the model. The count data included 192 rows (8 texts × 24 subjects) specifying the number of observed correct responses (max = 10 in each cell). The following syntax was used in *lme4* package [15]:

```
glmer(nCorr ~ Condition + (1+Condition|Subject) +
(1+Condition|Text), family = poisson, data = data)
```

Comparison of the full model to the reduced model (without the CONDITION predictor) led to statistically significant differences in model fit using likelihood-ratio tests ($\chi^2(1) = 4.94$, $p = 0.026$). The model predicts 1.85 (SE = 0.277) words recalled per text and subject in the *NoPause* condition but 2.41 in the *Pause* condition (SE = 0.370). However, singular fit was returned, indicating an excessively complex structure for the data. Therefore, we removed the by-text and by-subject slopes from the model; the intercept-only model produced almost identical predictions (*NoPause* = 1.85, SE = 0.28; *Pause* = 2.42, 0.359), and CONDITION was a significant predictor ($\chi^2(1) = 8.23$, $p = 0.004$). In conclusion, the effect – an advantage for *Pause* over *NoPause* of approximately 0.5 words per text – holds across subjects and texts.

For each word, we plotted its predictability from the pre-test and its recall in the experiment (Fig. 3a). Although the results varied (especially for zero predictability), there was a clear link between the two variables. The easier it was to guess the word, the better it was recalled after listening, producing a significant correlation ($r = 0.58$, $p < 0.001$). However,

Figure 3b plots the predictability vs. the effect size (*NoPause* condition subtracted from *Pause* condition; positive values mean the effect is in the expected direction). The absence of any significant correlation ($r = 0.15$, $p = 0.172$) confirms that the effect is due to the memory recall experiment manipulations.

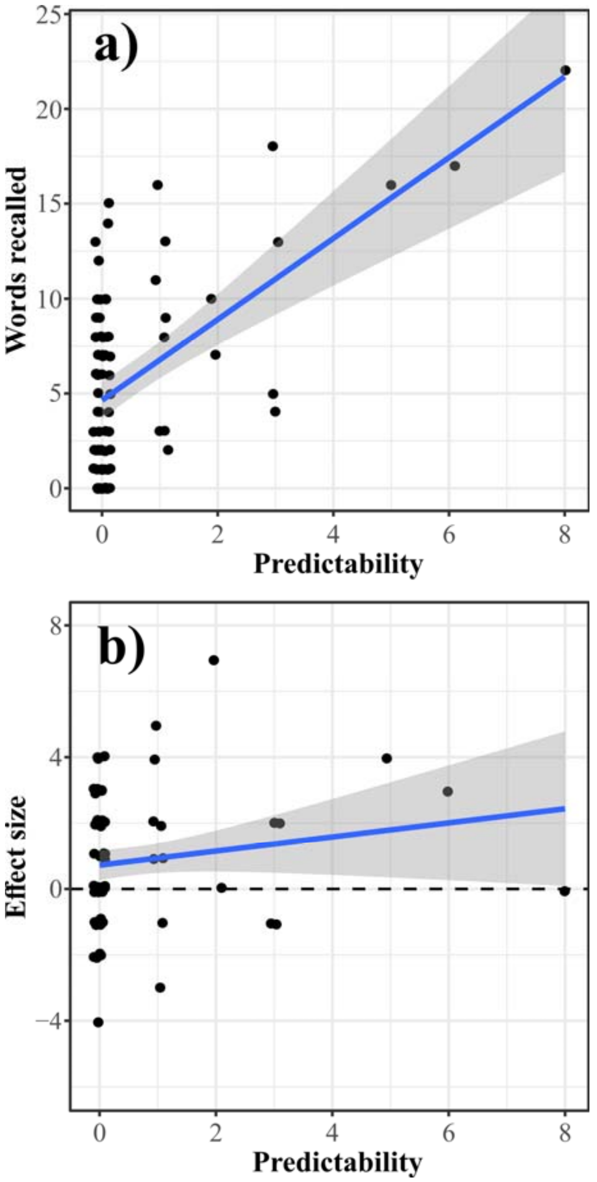

Figure 3: The relationship between predictability of words based on a pre-test and (a) recall in the experiment; (b) manipulation effect size. Jitter was applied to increase visibility of overlaid points.

4. DISCUSSION

The effect that we found may look impressive if we cite 265 correctly recalled items from texts with pauses against 198 items from the same texts without pauses. In reality, the effect is not particularly robust, even though it is statistically significant. It must be noted, however, that the more robust results reported in literature are usually achieved with highly artificial material that is distant from typical communicative language use. We, on the other hand, used narrative texts that were produced by renowned authors and told by experienced storytellers. The contents of the texts were not trimmed in any way, so their communicative purpose is genuine.

The exact underlying cerebral mechanisms responsible for the effect are still unclear. The authors in [6] speak about items that are "processed more deeply by listeners, resulting in improved recall". Such "deep processing" could probably be phrased as "more thorough processing that requires specified time lapse". If the time is unavailable (the speech keeps running), the processing is somehow incomplete. One fact from the past research that has been also corroborated by our experiment, however, seems to be apparent: listeners' memory is influenced by the prosodic structure of speech.

Our future work needs to turn to other prosodic cues as well. Researchers in [4] clearly showed that it is not "any grouping" that is beneficial for speech processing. In their experiment, they compared groupings delimited only by pauses with groupings delimited by clear intonation contours. Groups with melodic characteristics led to better memorability. Similarly, the influence of phrase-final deceleration, together with amplitude and voice quality changes, need to be examined.

It should also be noted that we only tested short-term memory retention. Helfrich and Weidenbecher found in their experiment that voice pitch does not affect immediate text retention, yet has a significant impact on long-term memory [12]. That poses still another research challenge.

Lastly, we are not aware of any study similar to ours on the Czech language. Ultimately, cross-linguistic comparisons in this field that take into account prosodic and syntactic specificities of Czech might be possible.

5. CONCLUSION

Pauses in spoken narratives should not be considered important only for speech production (breathing, neurolinguistics planning) but also for speech perception. The listeners in our experiment clearly benefited from their presence in a recall task. With the pauses absent in otherwise identical spoken texts, the lexical recall was significantly worse. Pauses thus clearly belong to the temporal structure of speech and fulfil important roles with regard to effectiveness of language communication.

6. ACKNOWLEDGEMENTS

The study was carried out with the support of GAČR (Czech Science Foundation), Project 21-14758S.

7. REFERENCES

[1] Baddeley, A. D. 1966. Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experim. Psychology* 18(4), 362–365.

[2] Turner, M. L., Engle, R. W. 1989. Is working memory capacity task dependent? *Journal of Memory and Language* 28, 127–154.

[3] Mathy, F., Feldman, J. 2012. What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition* 122, 346–362.

[4] Savino, M., Bosco, A., Grice, M. 2013. Intonation and positional effects in spoken serial recall. *Proc. of the Annual Meeting of the Cognitive Science Society*, 35, 3360–3365.

[5] Elmers, M., Werner, R., Muhlack, B., Möbius, B., Trouvain, J. 2021. Evaluating the effect of pauses on number recollection in synthesized speech. *Proc. 32nd Conference Elektronische Sprachsignalverarbeitung (ESSV '21)* Berlin, 298–295.

[6] Kimball, A. E., Yiu, L. K., Watson, D. G. 2020. Word recall is affected by surrounding metrical context. *Language, Cognition and Neuroscience,* 35(3), 383–392.

[7] Firbas, J. 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge University Press.

[8] Sanderman, A. A., & Collier, R. 1997. Prosodic phrasing and comprehension. *Language and Speech* 40, 391–409.

[9] Cutler, A., Dahan, D., van Donselaar, W. 1997. Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40, 141–201.

[10] Clifton, C., Carlson, K., Frazier, L. 2002. Informative prosodic boundaries. *Language and Speech* 45, 87–114.

[11] Dahan, D., Ferreira, F. 2019. Language comprehension: Insights from research on spoken language. In: Hagoort, P. (ed), *Human Language: From Genes and Brains to Behavior*, pp. 21–33. MIT Press.

[12] Helfrich, H., Weidenbecher, P. 2011. Impact of voice pitch on text memory *Swiss Journal of Psychology* 70 (2), 85–93.

[13] Hawkins, P.R. 1971. The syntactic location of hesitation pauses. *Language and Speech* 14, 277–288.

[14] Shipstead, Z., Lindsey, D. R., Marshall, R. L., Engle, R. W. 2014. The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language* 72, 116–141.

[15] Bates, D., Maechler, M., Bolker, B. & Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 526 67, 1–48.