PAUSE PARTICLES INFLUENCING RECOLLECTION IN LECTURES

Mikey Elmers

Language Science and Technology, Saarland University, Saarbrücken (Germany)
elmers@lst.uni-saarland.de

ABSTRACT

This study investigated the influence of pause-internal phonetic particles (PINTs) on recall for native and non-native listeners of English. Participants were 45 monolingual English and 45 L1 German listeners who heard segments from university lectures, in English, and answered content-based questions. Three versions of lecture stimuli were created: an unmanipulated original version, a "silence" version, and a "no PINTs" version where all PINTs were removed including silences. In the original and "silence" versions, half of the key information was preceded by PINTs material. The results indicated that material preceded by PINTs was less likely to be recalled. Additionally, the participant's first language was not significant for understanding the speaker. However, English listeners tended to score higher during the "no PINTs" condition, while German listeners tended to score higher during the original condition. This study was unable to replicate the recall benefit of PINTs found in single sentence laboratory setting experiments.

Keywords: pause particles, breath noises, memory, lectures, second language education

1. INTRODUCTION

Pause-internal phonetic particles (PINTs) encompass a variety of phenomena such as acoustic-phonetic silence, breath noises (i.e., exhalations and inhalations), filler particles (FPs) like "uh" or "um", and tongue clicks. These phonetic particles can exhibit an influence on recollection. For example, [1] found that the recollection of story plot points was improved when including FPs. In word recognition studies, [2] found that disfluencies improved the recollection of the following word, while [3] found that silent pauses improved the recollection of the following word. Importantly, [3] claims that a feature of disfluencies is that they provide additional time. [4] found that native and non-native listeners exhibited shorter response times for complex phrases that were preceded by FPs or silence compared to a no pause condition. Overall, these studies show that PINTs can affect recollection in laboratory settings. However, these studies do not utilize material from a real-world setting and focus on smaller contexts (i.e., words or sentences). This study expected to find a PINTs benefit for recollection in university lecture segments, similar to the previously mentioned smaller contexts.

Similar to [5], this study does not advocate for 'lab speech' or 'natural speech', rather the goal is to improve awareness around the types of data and methods used. This study explored the influence of PINTs on memory, using real-world data, rather than in a laboratory setting and with material larger than a single sentence. Another main goal was to evaluate the effect of PINTs on both native speakers (NSs) and non-native speakers (NNSs). I opted to examine English monolingual listeners and L1 German listeners due to the English language stimuli used in the study.

2. METHOD

Lectures were collected from Open Yale Courses [6] which contains free and open access courses from Yale University. English-language lectures were chosen based on the speaker's PINTs profile. After selecting a specific speaker, annotations were made for a subset of their lectures. The chosen speaker displayed a relatively high number of PINTs during his lectures, with upwards of 40% of his total time incorporating PINTs material. Fig. 1 shows an example for this speaker.

2.1. Participants

This study used a web-based experiment created with Labvanced [7] to present the audio stimuli to participants, and to collect their answers and questionnaire information. Participants were recruited using the crowd-sourcing platform Prolific [8] and consisted of 45 monolingual English participants (mean age 38 years; age range 21–62 years) and 45 L1 German participants (mean age 35 years; age range 21–72 years) who were paid for their participation. One monolingual English participant reported hearing impairment and was not included in the results.

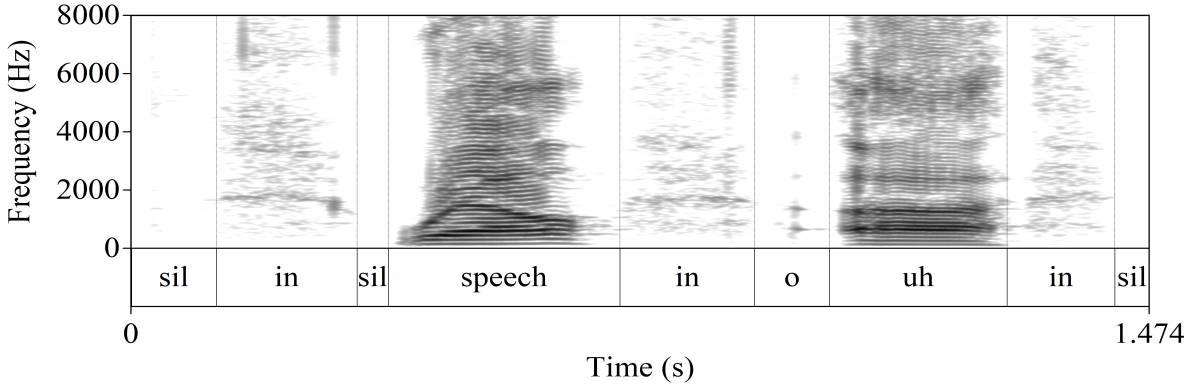

Figure 1: Example section from speaker. Annotations of PINTs: silence (sil), inhalation noise (in), exhalation noise (ex), filler particles uh and um, tongue click (cl), and other (o).

2.2. Stimuli

Stimuli consisted of four three-minute sections extracted from full length lectures. Each audio segment was followed by two multiple-choice content-based questions, with one question preceded by PINTs material and the other not. The study was balanced so that the key material was equally preceded, or *not* preceded, by PINTs. However, neither question was preceded by PINTs material in the "no PINTs" condition, since all PINTs material was removed. An example question was: "According to Paul Fussell, what is the essential trope or rhetorical figure of World War One poetry?" The possible answers were: a) hyperbole, b) metaphor, c) oxymoron, d) irony. The participants did not need to know what these concepts meant, or any other encyclopedic or background knowledge. Instead, they needed to answer based on the content as presented by the lecturer.

The different conditions were created using a Praat [9] script that would remove or replace the PINTs material. In the "silence" condition, non-silence PINTs were replaced with a silence taken from the audio and matched to the duration of the cut material. Therefore, the "silence" condition maintained the same duration as the original audio. The original and "silence" conditions provided the same amount of processing time, while the "no PINTs" condition provided less processing time (see Fig. 2). The "no PINTs" condition did not include any acoustic pause whatsoever. Participants only heard one of the three conditions, i.e., one third of participants heard four original clips, one third of participants heard four "silence" clips, and one third of participants heard four "no PINTs" clips. Each of the conditions included the same textual material, however, the order of the four audio clips was randomized to prevent ordering effects.

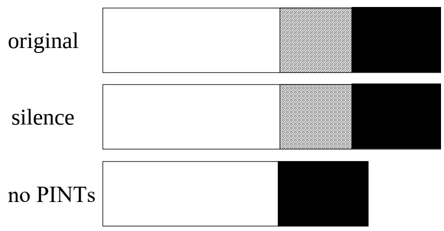

Figure 2: Schematic of the duration for the three conditions showing speech (white), PINTs (grey), and speech material that contained the key information (black).

2.3. Procedure

Participants were informed that they would hear four audio clips, each approximately three minutes, and answer content-based questions immediately following each clip. Participants were instructed to use headphones and test their audio volume before starting. They were told to not take notes. They were also told that the recordings were from a non-ideal microphone and included some background noise. This was in order to draw their attention away from some of the minor artefacts that occurred from the audio manipulation in the "silence" and "no PINTs" conditions. Participants were told that they would receive a score at the end of the test as an additional incentive to perform well. While listening to the audio, participants saw "Listen closely!" on their screen. They heard each audio clip only once.

After completing the listening section, participants answered a questionnaire that included: age, hearing impairment, L1, self-assessed English skills (for the German listeners), a test score if possible (for the German listeners), highest completed education (high school, university, or other), level of interest in the audio contents (1: very uninterested to 5: very interested), how easy the speaker was to follow and understand (1: very difficult to 5: very easy), and how prepared they found the speaker (1: very unprepared to 5: very prepared). Total completion time was between 15-20 minutes.

3. RESULTS

Participants were scored based on how many questions they answered correctly with a maximum score of 8 (1 point per correct answer). The monolingual English participants scored higher than the L1 German participants in all conditions except in the original condition, however, the monolingual English speakers usually had a higher variance (see Table 1). Monolingual English participants scored highest during the "no PINTs" condition while the L1 German participants scored highest during the original audio condition.

condition	L1	mean	median	sd	n
noPINTs	*EN*	6.26	7	1.83	15
silence	*EN*	6.07	6	1.44	13
original	*DE*	6.00	6	1.07	15
original	*EN*	5.88	6	1.92	16
noPINTs	*DE*	5.87	6	1.41	15
silence	*DE*	5.67	5	1.76	15

Table 1: Mean score, median score, standard deviation score, and count information for the different conditions and L1s.

This project's data and scripts can be found at https://github.com/MikeyElmers/paper_icphs23. The data was pre-processed using the dplyr [10] (Version 1.1.1), stringr [11] (Version 1.5.0), and tidyr [12] (Version 1.3.0) packages. Homogeneity of variance was evaluated using Levene's test from the car [13] (Version 3.1.2) package. Statistical models were analyzed with linear regression and binomial generalized linear mixed models (binomial GLMMs) with the lme4 [14] (Version 1.1.31) and lmerTest [15] (Version 3.1.3) packages in R [16] (Version 4.0.4). Models were compared with the Akaike information criterion (AIC) [17] to calculate unexplained variance. The best fit model was selected as the model with the lowest AIC.

3.1. Preceding PINTs

This study investigated the effect of PINTs material immediately before key information on participant score. These models did not include the "no PINTs" audio condition since all PINTs material was removed. Scores are out of 1 rather than 8 since the evaluation is done on a by-question basis rather than a subject's collective score. Table 2 shows that when key information was preceded by PINTs, the result was an overall lower score. The data showed violations for normality, as indicated by the Shapiro-Wilk test, and homogeneity of variances, as indicated by Levene's test. Therefore, the non-parametric Wilcoxon rank sum test was used. Results indicated a significant difference between the preceding PINTs conditions ($W = 32096$, $p < 0.001$). Participants performed significantly better when critical information was *not* preceded by PINTs information.

preceding PINTs	mean	sd
no	0.81	0.39
yes	0.66	0.47

Table 2: Mean and standard deviation score based on whether the answer was immediately preceded by PINTs material.

Binomial GLMMs were used to evaluate which variables influenced score. The model with the best fit was: $glmer(score \sim precede + (1 \mid id), family = binomial)$. This model predicts score based on the answer being preceded by PINTs information as a fixed effect, and subject with intercept as a random effect. This model performed better than models that incorporated L1, condition, or the questionnaire variables. The analysis revealed a main effect for preceding PINTs ($Estimate = -0.88$, $SE = 0.23$, $z = -3.87$, $p < 0.001$). This main effect indicates that the presence of PINTs material before the answer lowered participants' score.

3.2. Ease

Participant's reported how easy it was to follow and understand the speaker (1: very difficult to 5: very easy). Overall, the mean ease was 2.82. The condition that removed all the pause material was considered the easiest to follow when averaging over all participants. However, an ANOVA showed no significant differences between conditions (*F(2, 86) = 0.88, p > 0.05*). This finding was interesting since substituting or deleting PINTs material created minor artefacts within the audio. The original, unmanipulated version was found to be the most difficult to follow, possibly because this speaker uses a high frequency of PINTs ($\sim$ 40% of his total speaking time). When comparing means of ease by L1, I found that the monolingual English group (M = 3.05 , SD = 1.29) and the L1 German group (M = 2.60, SD = 1.16) were not significantly different (*t(87) = 1.71, p > 0.05, d = 0.36*). These results indicate that the NNSs found the English-language lecturer as easy to understand as the monolingual NSs of English.

Linear regression models were tested with L1, condition, and the different questionnaire variables. The model with the best fit (lowest AIC) predicted total score with ease as the only fixed effect. Table 3 shows that with an ease value of 1, participants' total score was 4.84 (out of a total of 8) and that the higher the ease value, the higher the total score. Significant effects for all levels of ease were found, except for a value of 3. Importantly, an ease value of 5 improved participants' total score more than an ease value of 4 which improved more than an ease value of 2. The ease value of 3 did not follow this trend.

	Est	SE	t	p-value
$(Intercept)$	4.84	0.42	11.60	$< 0.001***$
$ease2$	1.15	0.51	2.28	$< 0.05*$
$ease3$	0.95	0.54	1.77	0.08
$ease4$	1.54	0.55	2.81	$< 0.01**$
$ease5$	1.95	0.64	3.09	$< 0.01**$

Table 3: Summary information of linear model with ease as predictor.

3.3. Correlation

Table 4 contains the Pearson correlations between total score and the questionnaire variables. No correlation was found between age or interest and total score. However, I found a weak correlation between ease and total score (*t(87) = 2.98, p < 0.01, r = 0.30*), and between preparation and total score (*t(87 = 2.77, p < 0.01, r = 0.28*).

4. DISCUSSION

This experiment evaluated whether PINTs improved the recall of English-language lecture material for native and non-native listeners. I found that PINTs immediately preceding key information negatively impacted score. While I found that L1 did not influence the ease rating of the speaker, non-native students may encounter significant problems when listening to lectures, such as word recognition or with creating meaning [18]. These issues are related to linguistic aptitude and awareness of the lecture material. This study found that monolingual English listeners tended to scored better in the "no PINTs" situation, while L1 German listeners tended to scored better with the original audio. This may be due to monolingual English listeners being above the threshold of needing the time-buying aspect of PINTs. Conversely, the L1 German listeners still benefited from the time-buying aspect of PINTs. [19] found that a base skill level might be required before the benefits of additional processing time from pauses can be seen, and importantly, that beyond a certain skill level pauses may no longer aid and, instead, be an irritant to the listener. [20] found that pauses may increase comprehension, but only for advanced students. [20] also found that increasing the duration of pauses has a ceiling and once above that ceiling, comprehension will decrease. Therefore, it is important to consider the impact of PINTs in environments where the recall of key information is crucial, such as educational settings, and for both native and non-native listeners.

	age	ease	interest	prep
total score	0.09	0.30	0.10	0.28

Table 4: Pearson correlations between total score and questionnaire responses for all participants.

This study investigated the influence of PINTs on recollection in an ecologically valid scenario with longer material lengths. I found PINTs to be detrimental to the recollection of upcoming content in a lecture scenario, which is contradictory to what previous studies have found. These results might differ from other studies due to the longer material lengths, or because there are many variables that are difficult to control in a real-world scenario. Additionally, this experiment treated all PINTs equally. In real-world scenarios, speakers have distinct PINTs profiles and often many PINTs will co-occur making it difficult, if not impossible, to evaluate individual PINTs separately. Additionally, while non-native participants were asked for their test score, I was not able to get an accurate picture of the influence of PINTs for different skill levels. In this study most of the L1 German listeners were advanced in English. Future work should continue to evaluate longer material lengths, with a variety of language backgrounds and skill levels.

5. REFERENCES

[1] S. H. Fraundorf and D. G. Watson, "The disfluent discourse: Effects of filled pauses on recall," *Journal of memory and language*, vol. 65, no. 2, pp. 161–175, 2011.

[2] M. Corley, L. J. MacGregor, and D. I. Donaldson, "It's the way that you, er, say it: Hesitations in speech affect language comprehension," *Cognition*, vol. 105, no. 3, pp. 658–668, 2007.

[3] L. J. MacGregor, M. Corley, and D. I. Donaldson, "Listening to the sound of silence: Disfluent silent pauses in speech have consequences for listeners," *Neuropsychologia*, vol. 48, no. 14, pp. 3982–3992, 2010.

[4] M. Watanabe, Y. Den, K. Hirose, and N. Minematsu, "The effects of filled pauses on native and non-native listeners' speech processing," in *Disfluency in Spontaneous Speech*, 2005.

[5] P. Wagner, J. Trouvain, and F. Zimmerer, "In defense of stylistic diversity in speech research," *Journal of Phonetics*, vol. 48, pp. 1–12, 2015.

[6] "Open yale courses," https://oyc.yale.edu/, accessed: June 6th, 2022.

[7] H. Finger, C. Goeke, D. Diekamp, K. Standvoß, and P. König, "Labvanced: a unified JavaScript framework for online studies," in *International Conference on Computational Social Science (Cologne)*, 2017.

[8] "Prolific," Oxford, UK, 2014, accessed: 12.06.2022. [Online]. Available: https://www.prolific.co

[9] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 2022. [Online]. Available: http://www.praat.org

[10] H. Wickham, R. François, L. Henry, K. Müller, and D. Vaughan, *dplyr: A Grammar of Data Manipulation*, 2023, r package version 1.1.1. [Online]. Available: https://CRAN.R-project.org/package=dplyr

[11] H. Wickham, *stringr: Simple, Consistent Wrappers for Common String Operations*, 2022, r package version 1.5.0. [Online]. Available: https://CRAN.R-project.org/package=stringr

[12] H. Wickham, D. Vaughan, and M. Girlich, *tidyr: Tidy Messy Data*, 2023, r package version 1.3.0. [Online]. Available: https://CRAN.R-project.org/package=tidyr

[13] J. Fox and S. Weisberg, *An R Companion to Applied Regression*, 3rd ed. Thousand Oaks CA: Sage, 2019. [Online]. Available: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

[14] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[15] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest package: Tests in linear mixed effects models," *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[16] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: https://www.R-project.org/

[17] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *International Symposium on Information Theory*, 1973, pp. 267–281.

[18] J. Kilbon, "L2 student perceptions regarding their comprehension of academic lectures–a longitudinal study," Ph.D. dissertation, University of Leicester, 2022.

[19] E. K. Blau, "The effect of syntax, speed, and pauses on listening comprehension," *TESOL quarterly*, vol. 24, no. 4, pp. 746–753, 1990.

[20] G. Jacobs, W. Chuawanlee, B. K. Itoga, D. Sakumoto, S. Saka, and K. A. Meehan, "The effect of pausing on listening comprehension," in *The Eighth Second Language Research Forum*. Honolulu, HI: ERIC, 1988, pp. 1–19.