

Mikey Elmers¹, Éva Székely²

¹Saarland University (Germany), ²KTH Royal Institute of Technology (Sweden)

elmers@lst.uni-saarland.de, szekely@kth.se

Background

- Pause-internal phonetic particles (PINTs) include silences, inhalation and exhalation noises, filler particles “uh” and “um”, and tongue clicks
- PINTs benefit in single-sentence laboratory setting experiments for text-to-speech (TTS) [1-3]
- Many studies do not utilize material from a real-world setting and/or focus on smaller contexts (i.e., words or sentences)

Research Questions: Do PINTs improve recall in lectures? Do PINTs affect recall differently for L1 and L2 listeners?

Method

- Generated stimuli with neural TTS [4]
- Three versions: base, silence, and no PINTs (Fig. 1)
- Half of key information preceded by PINTs material
- Half of participants told they were listening to computer-generated audio (cg)
- 90 L1 English (monolingual) and 90 L1 German participants
- Participants heard 4 lecture segments (3-minutes each)
- Participants answered 2 content-based questions

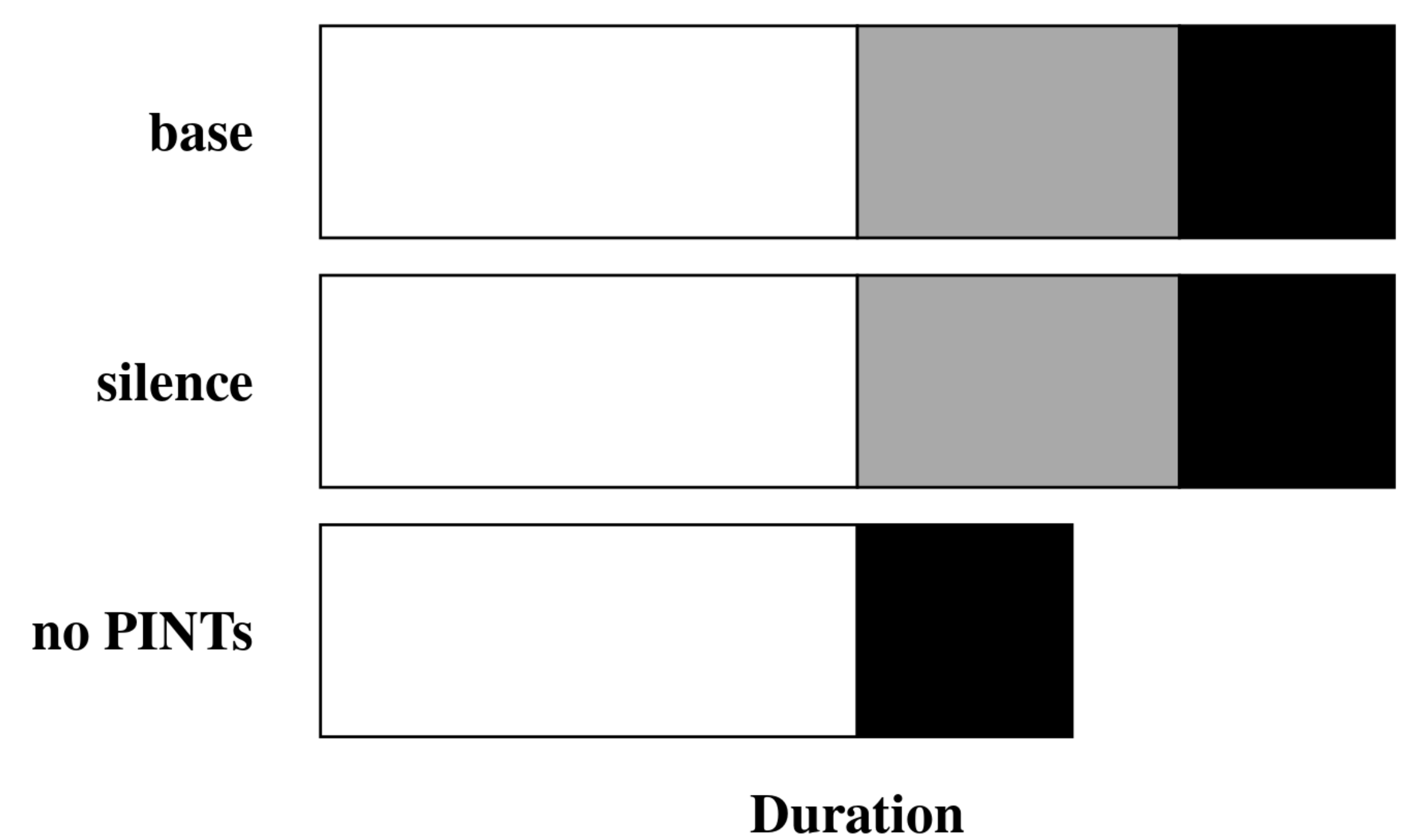


Fig. 1: Schematic for three conditions: speech (white), PINTs (grey), and speech material containing key information (black).

Results

- Participants scored 0-8 (1 point per question) (Fig. 2)
- Instruction type did not affect recall
- Omitted no PINTs condition for modeling
- Binomial GLMM model:
 - $glmer(score \sim precede + interest + (1|id), family = binomial)$
- Main effect for preceding PINTs:
 - $\beta = -0.96, p < 0.001$
- Key information preceded by PINTs lowered score (Fig. 3)

condition	mean	median	sd
silence _{cg}	5.77	5	1.41
nopints	5.71	6	1.61
base _{cg}	5.61	6	1.64
base	5.45	5	1.78
nopints _{cg}	5.34	5	1.72
silence	4.77	5	1.68

Fig. 2: Descriptive statistics for total score for the different conditions.

preceding PINTs	mean	sd
no	0.77	0.42
yes	0.58	0.49

Fig. 3: Descriptive statistics for by-question score. Wilcoxon rank sum test ($W = 132396, p < 0.001$).

Summary

- Material preceded by PINTs less likely to be recalled
- L1 did not affect recall
- Unable to replicate recall benefit found in single-sentence laboratory settings

References

- [1] Dall, R., Tomalin, M., & Wester, M. 2016. Synthesising filled pauses: Representation and datamixing. In Proc. 9th ISCA Workshop on Speech Synthesis (SSW 9). pp. 7–13. [2] Elmers, M., Werner, R., Muhlack, B., Möbius, B., & Trouvain, J. 2021. Evaluating the effect of pauses on number recollection in synthesized speech. In Proc. 32nd Conference Elektronische Sprachsignalverarbeitung (ESSV '21). pp. 289–295. [3] Elmers, M., Werner, R., Muhlack, B., Möbius, B., & Trouvain, J. 2021. Take a breath: Respiratory sounds improve recollection in synthetic speech. In Proc. Interspeech 2021. pp. 3196–3200. [4] Elmers, M., O'Mahony, J., and Székely, É. 2023. Synthesis after a couple PINTs: Investigating the role of pause-internal phonetic particles in speech synthesis and perception. In Proc. Interspeech 2023 (Accepted).