

# Pause-internal particles in university lectures

Mikey Elmers, Jürgen Trouvain

*Language Science and Technology, Saarland University, Saarbrücken, Germany*

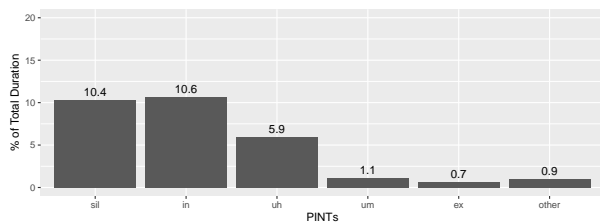
Pauses consist of one or more pause-internal particles (PINTs) such as: silences, breath noises, filler particles, and tongue clicks. In speech technology and phonetics, PINTs are often overlooked. This is surprising because there is a strong need for exploring expression styles in these fields, and PINTs can certainly help. These particles provide a great deal of information, and despite their relatively short duration, they exert a large influence upon style [1], perception [2], and memory [3].

University lectures showcase unique, style-specific features. For example, lectures are semi-prepared and somewhat-rehearsed while simultaneously spontaneous and monologic. This is in contrast to sentences recorded in a laboratory setting with a neutral style. Regarding lectures Kjellmer [4, p.190] stated, "a lecture that is read aloud from the written page is often difficult to take in when its delivery lacks the verbal guides and signposts that we more or less subconsciously expect to find in speech; as listeners we are in danger of missing the point of the argument."

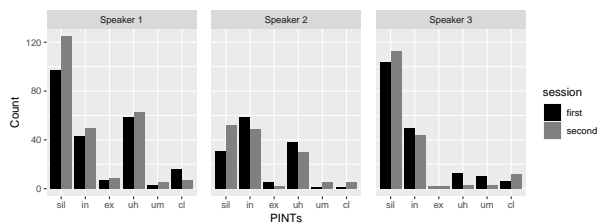
In this pilot study, recorded university lectures were annotated for the usage of different PINTs. The lectures were chosen from Open Yale Courses [5]. We selected 3 speakers and annotated 6 minutes for each speaker. The audio was annotated with one of the following labels: *speech*, *silence*, *inhalation*, *exhalation*, *uh*, *um*, *tongue click*, or *other* (e.g. coughing, swallowing, etc.). No minimum threshold was enforced for any of the PINTs. Annotations included 3 minutes from the beginning of the semester and 3 minutes from the end of the semester, to evaluate speaker consistency. The annotations were compiled for duration, count, and frequency information. We found over 1100 PINTs in the 18 minutes of annotated material. Approximately 30% of the total time consisted of one of the investigated PINTs (Fig. 1a). Both silences and inhalations each accounted for approximately 10% of the total time, with the remaining PINTs accounting for the last 10%. Each lecturer displayed a particular pattern regarding the type and number of PINTs used compared to the other lecturers (Fig. 1b). Despite the differences in the amount and type, the duration of each PINT type was consistent across speakers. Additionally, each lecturer tended to be consistent with their PINTs usage throughout the semester. Speakers also varied in their PINTs frequency, with some having a PINTs rate as low as 0.75 PINTs/sec, and others with twice as many.

Since teaching time is valuable, we expected to find few, if any, PINTs. Instead we found many particles that accounted for more than 30% of the total time. This finding is in stark contrast with current synthesis techniques which include silences but omit the other PINTs. Meaning, current synthesis techniques ignore about 20% of material, at least for this speech genre. Additionally, these results are evidence against the belief that silence dominates pauses since approximately 2/3 of the entire pause duration consists of particles that are not silent.

This work can function as a baseline for achieving a more natural usage of PINTs for lectures in speech synthesis. Another notable application is for language learning and testing. For example, the listening components on language proficiency tests, like the TOEFL, often include lectures but with minimal PINTs. It is important to investigate whether these particles have an influence on the recollection of lecture material, and if the effects are consistent for both native and non-native speakers.



(a)



(b)

Figure 1: (a) includes the total duration of each PINT type as a percent. Speech accounts for  $\sim 70\%$  and the PINTs account for  $\sim 30\%$  of the total duration. (b) displays the number of each PINT for each speaker with color indicating first/last day of semester. Labels: *silence* (sil), *inhalation* (in), *filler particle* (uh), *filler particle* (um), *exhalation* (ex), and *other* particles (e.g., tongue click, coughing, swallowing, etc.).

## References

- [1] Jürgen Trouvain and William J Barry. The prosody of excitement in horse race commentaries. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, pages 86–91, 2000.
- [2] Hans Rutger Bosker, Hugo Quené, Ted Sanders, and Nivja H De Jong. The perception of fluency in native and nonnative speech. *Language Learning*, 64(3):579–614, 2014.
- [3] Martin Corley, Lucy J. MacGregor, and David I. Donaldson. It’s the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, 105:658–668, 2007.
- [4] Göran Kjellmer. Hesitation. in defence of er and erm. *English Studies*, 84(2):170–198, 2003.
- [5] Open yale courses. <https://oyc.yale.edu/>. Accessed: June 6th, 2022.