

Pertsch: A Corpus of Persian and German Based on Different Speech Elicitation Tasks

Creator

Neda Mousavi (Martin Luther University Halle-Wittenberg)
ORCID: 0009-0006-1903-3971

Citation

Mousavi, Neda. (2026), Pertsch: A Corpus of Persian and German Based on Different Speech Elicitation Tasks, Preprint DOI: [https://doi.org/ 10.31234/osf.io/ya3c5_v2](https://doi.org/10.31234/osf.io/ya3c5_v2)

Description

The Pertsch Corpus is a bilingual speech corpus designed to investigate speech production across different communicative situations in Persian and German. The corpus contains recordings from 60 adult speakers (30 Persian, 30 German; balanced for gender), each completing a fixed sequence of seven speech elicitation tasks ranging from controlled read speech to more open-ended communicative conditions.

The elicitation tasks include:

- formal read speech
- informal read speech
- spontaneous conversation
- picture description
- autobiographical storytelling
- informal voice messages
- formal voice messages

All recordings were collected under standardized laboratory conditions and are provided as high-quality mono WAV files (44.1 kHz, 16-bit). The corpus includes:

- audio recordings
- orthographic transcriptions
- Praat TextGrid annotation files
- phonetic segmentation
- consonant/vowel interval tiers
- vowel onset tiers
- annotation layers for pauses, fillers, repairs, and speech intervals

The corpus includes task-based laboratory speech recordings with orthographic transcription, automatic alignment using WebMAUS, generation of consonant and vowel interval annotations using Praat scripts, and manual annotation of verbal and non-verbal speech elements.

The corpus was originally developed as part of the PhD dissertation “Individual speech rhythm in Persian and German: An acoustic–cognitive approach of rhythm production in different speaking tasks” by Neda Mousavi (2025), supervised by Prof. Dr. Sven Grawunder.

Languages

Persian (fa)

German (de)

Data Collection Date

2024

File Formats

WAV

TextGrid

TXT / metadata tables

Acknowledgments

The Pertsch corpus was developed as part of the author’s PhD dissertation in Speech Science, entitled “Individual Speech Rhythm in Persian and German: An Acoustic--Cognitive Approach of Rhythm Production in Different Speaking Tasks”, supervised by Prof. Dr. Sven Grawunder. The author gratefully acknowledges the financial support received during the PhD studies from the Goethe Research Academy (GRADE), the Soudavar Memorial Foundation, and Martin Luther University Halle-Wittenberg.

Related Identifier

OSF subset release:

<https://doi.org/10.17605/OSF.IO/4S3GZ>