# ASSESSING REGISTER VARIATION IN LOCAL SPEECH RATE

Daniel Duran[1], Melanie Weirich[2] & Stefanie Jannedy[1]

[1]Leibniz-Centre General Linguistics (ZAS), Germany [2]Friedrich Schiller University Jena, Germany
duran@leibniz-zas.de, melanie.weirich@uni-jena.de, jannedy@leibniz-zas.de

## ABSTRACT

We present results on phonetic register variation (i.e. conventionalized and socially recurring linguistic patterns of intra-individual speech behavior) in contrasting situated interactions. In this paper, we analyze intra-individual differences in local speech rate. Data was elicited with a novel method where participants talk to a video-taped interlocutor in a simulated tele-conference. Each participant mastered a "formal" situation (e.g. requesting a deadline extension from a superior) and an "informal" one (e.g. describing their favorite recipe to an acquaintance). The silent interlocutor persona appeared in two guises differing in hair style, clothing and make-up. This set-up allows us to systematically and consistently manipulate the experimental condition while eliciting laboratory induced differences in fine phonetic detail. We hypothesize that formal situational requirements slow down speech production due to processing and memory load constraints. Statistical analyses for 45 German participants show a slower speech rate in the formal condition.

**Keywords**: Speech register, intra-speaker variation, formality, speech rate, sociophonetics.

## 1. INTRODUCTION

This work is embedded in elucidating the various levels of linguistic awareness that speakers have when making contextual linguistic choices. Such selections are being made because speakers do not act uniformly, independent of context, situation or interlocutor. Rather, speakers seem to adjust their vocabulary and wording, forms of addressing and most certainly also the fine phonetic detail of their speech.

In fact, a social awareness of variation in speech style depending, for example, on the addressee [1] or referee [2] has been shown previously. Specific studies on cross-situational variation of single speakers showed that differences in fine phonetic detail like the use of phonation type [3], creak [4] or palatalization of fricatives [5] may serve to construct differences in social meaning, group affiliation and personas. Differences in speech style due to situational (i.e. level of formality or in Labov's sense "attention paid to speech" [6]) and functional (i.e. request, narration) variation and perceived social rank of interlocutor we here refer to as differences in register [7, 8]. Previous studies found a slower speech rate in polite and formal speech in contrast to informal speech in German or Dutch [9, 10]. A slower speech rate has also been associated with a higher cognitive load [11, 12], though findings of effects in the opposite direction, i.e. faster speech rates under higher cognitive load, have also been reported [13].

In order to study if and how fine phonetic detail is being altered based on the speech situation, the level of formality and the interlocutor, we have devised an experimental paradigm that tightly controls for these three factors. The paradigm is novel in the sense that it simulates a set-up that people have become quite used to during the pandemic. Pre-recording the addressee assured that all facettes of the set-up were held constant: the movements, gaze and posture of the interlocutor as well as their clothing, makeup and hair style and the surrounding prompts (coffee cup, smart phone). Thus, in our work, we set out to elicit laboratory induced intra-individual stylistic differences in fine phonetic detail, specifically, in speech rate, with the interlocutor held constant. Our hypothesis is that formal situational requirements slow down speech production due to processing and memory load constraints while casual and informal conversations facilitate faster speaking rates.

## 2. METHOD

### 2.1. Stimuli

In a pre-study, the level of formality of the interlocutor was rated by 26 participants on 15 personality attributes. Two stimulus videos were produced based on these previous ratings of perceived personality traits given different hair styles, clothing and make-up of the depicted interlocutor. One video was produced to be perceived as most formal and one to be perceived as least formal. The silent interlocutor persona thus appeared in two guises differing in hair style, clothing and make-up in the formal and the informal stimulus videos (cf. Figure 1). A third neutral video was produced to serve as a familiarization and test-trial during the experiments. One female actor was recorded sitting at a desk and changing slightly in head, arm and upper body position in a choreographed manner in these different guises while apparently listening to her interlocutor but not speaking.

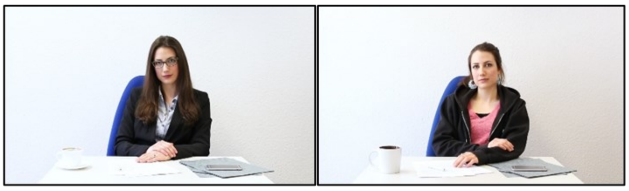

**Figure 1:** The two personas chosen for portraying the greatest (left) and the least (right) level of formality.

### 2.2. Task

The experimental setup was designed to simulate a video-call. Participants were seated at a table in front of a large screen which showed the stimulus video. The duration of each stimulus video was roughly 2 minutes. Participants were informed of the duration of each recording and asked to fill the entire recording time if possible. To facilitate conversation, ideas for things to mention were presented to each participant. After the familiarization scenario, participants were recorded in two conditions: a formal scenario and an informal scenario. Three cover stories were prepared for each of the two settings and each participant was assigned one story for the formal and one story for the informal setting.

The participants' task was to talk to the pre-recorded interlocutor in two situations: (a) in the formal, "at-stake" situation, for example, they had to request a deadline extension for a term paper or negotiate a pay raise with a superior, and (b) in the informal, "nothing-at-stake" situation, they had to describe their favorite recipe or their favorite things to do around town to a new neighbor or an acquaintance. These situations were designed to induce the largest possible contrast between the social contexts in terms of formality. Each participant interacted with the formal and the informal video in randomized order.

### 2.3. Participants

A total of 45 participants were recorded: 15 speakers were recruited in Bremen (9 female, 6 male) and 30 speakers in Berlin (16 female, 14 male, mean age: 24, range: 19 to 35 years). All participants received written information about the procedure and data handling and gave their written consent prior to each recording session. A compensation in accordance with the German minimal-wage requirement was paid for each participation. None of the participants included in this analysis reported any known speech impairments. They were all German speaking residents of Bremen or Berlin (and surroundings).

Data was recorded in two different environments: the group of speakers from Bremen were recorded in a home environment (due to contact restrictions during the early phases of the pandemic, "home space") and speakers from Berlin have been recorded in a sound-attenuated laboratory setting ("lab space").

### 2.4. Speech Material

All speech data was segmented and annotated with WebMAUS [14] with the "German (DE)" language model. All annotations were manually corrected in Praat [15]. Syllables have been generated automatically with the "emuR" R package [16]. Canonical transcriptions have been adopted from MAUS with manual corrections according to [17], where necessary.

We evaluated approx. 165 minutes of speech data from the 45 speakers. Excluding pauses, approximately, 117 minutes of total vocalization time were analyzed. Figure 2 visualizes the amount of speech data by speaker gender (female vs. male), experimental condition (formal vs. informal) and space (Berlin lab vs. Bremen home). On average, male participants tend to speak less than female participants. Also, both participant groups, tend to speak more in the informal condition in comparison to the formal condition.

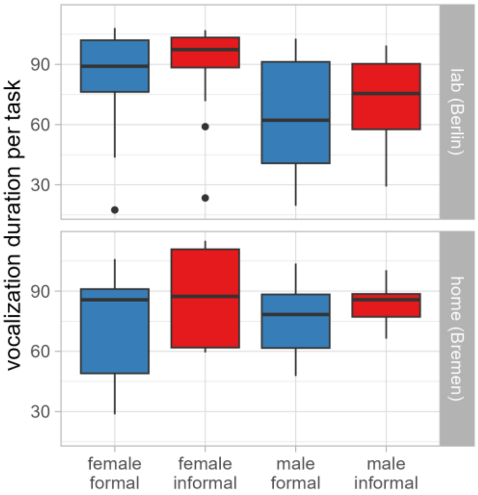

**Figure 2**: Speech material (only words excluding pauses, i.e. total vocalization time) by speaker gender, condition and recording environment.

The local speech rate was assessed for a total of 7,531 inter-pausal intervals as the number of syllables per second (SPS), calculated as the number of syllables divided by the total duration of the syllables measured in seconds. In addition, we compute "reduction" as the percentage of realized phonetic segments from the total number of canonical segments (based on the manually corrected canonical transcriptions of each word segment). In other words, we compute reduction as the proportion of "realized" sounds to "intended" phonological segments [cf. 18].

## 3. ANALYSIS AND RESULTS

We fit linear mixed-effects models (LMER) in R using the "lme4" package [19] for the dependent variables *total vocalization duration (TVD)*, *syllable rate* (expressed as syllables per second, *SPS*) and *reduction*. Speaker *gender* ("female" vs. "male"), *space* ("Bremen home" vs. "Berlin lab") and *condition* ("informal" vs. "formal") are defined as categorical fixed effects. We also include the *session duration* for the modelling of total vocalization time as a control variable. For easier interpretation, we center the values for session duration, which sets the reference level at the intercept of the model to the mean duration of all sessions. We include *speaker* as random effect. We apply forward selection in model fitting by step-wise adding variables and all two- and three-way interaction terms and comparing model fits based on ANOVAs. We compute post hoc pairwise comparisons using the "emmeans" package [20].

### 4.1 Total vocalization duration (TVD)

We first model total vocalization duration (*TVD*) in order to identify differences between sub-groups within our data set. We start with an intercept-only null model including only random intercepts by speaker. Adding *condition* improves the model fit ($\chi^2(1) = 6.474$, $p = 0.011$). The variables *gender* or *space* do not further improve the model fit. However, adding the interaction term *condition*space* improves the model ($\chi^2(2) = 6.523$, $p = 0.038$). We also find that adding the centered session duration (*CSD*) improves the model ($\chi^2(1) = 53.763$, $p < 0.001$). The best fitting model for the total vocalization time is (in R notation):

$$TVD \sim CSD + condition*space + (1|speaker)$$

The intercept estimate is at 73.98 (SE = 3.54). This corresponds to the base levels *condition* = "formal", *space* = "Berlin lab", and *CSD* = 0, i.e. the mean session duration. The term for session duration is significant (est. 0.793, SE = 0.087, $p < 0.001$), indicating as expected more vocalization in longer session durations. Pairwise comparisons between "formal" and "informal" by *space* show no significant difference for the Berlin-lab group ($p = 0.569$) or the Bremen-home group ($p = 0.881$). Note though that the estimate is larger for the lab speakers (est. 1.385, SE = 2.59) than for the home speakers (est. 0.555, SE = 3.70). Comparisons between the "lab" and "home" *space* settings by formality *condition* show a significant difference between the Berlin-lab and the Bremen-home participants in formal ($p = 0.018$) and informal ($p = 0.023$) conditions, indicating longer vocalization durations in the home space independent of condition. Figure 3 shows the data by *condition* and *space*, and the effect plot of the interaction between these variables.

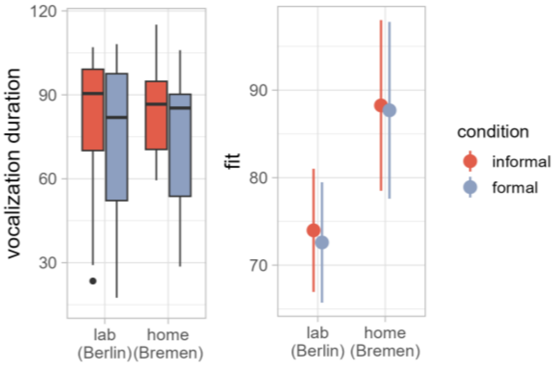

**Figure 3**: Vocalization duration in seconds by space and condition (left) and the predicted model fit (right panel).

### 4.2 Local speech rate in SPS

We again start with an intercept-only model for the local speech rate in SPS. Adding *condition* improves the model fit ($\chi^2(1) = 6.149$, $p = 0.013$). Adding *gender*, *space* or any interaction term does not further improve the model. The best fitting SPS model is thus:

$$SPS \sim condition + (1|speaker)$$

The intercept for *condition* = "informal" is at 4.78 (SE = 0.09). The *condition* term is statistically significant (est. -0.137, SE = 0.055, $p = 0.013$), i.e. the model predicts fewer SPS in the formal condition in comparison to the informal condition. Figure 4 shows the local speech rate in syllables per second (SPS) by experimental condition and the model fit.

### 4.3. Reduction

We start building the reduction model with an intercept-only model including only random intercepts by speaker, as with the previous two models. Adding *gender* ($\chi^2(1) = 4.824$, p = 0.028), *space* ($\chi^2(1) = 20.035$, p < 0.001) and their interaction *gender*space* ($\chi^2(1) = 5.169$, $p = 0.023$) improves the model fit. Adding *condition*, however, does not improve the model fit ($\chi^2(1) = 3.522$, $p = 0.061$), nor does any interaction with condition. The best reduction model is thus:

$$reduction \sim gender*space + (1|speaker)$$

The intercept of the reduction model is at 11.83 (SE = 0.81), i.e. for *gender* = "female" and *space* = "Berlin lab". Pairwise comparisons between "male" and "female" by *space* show a significant gender difference in reduction for the "lab" participants ($p$ =

0.002), but not for the “home” participants ($p$ = 0.684). A pairwise comparison of *space* by *gender* shows significant differences between home and lab in phonetic reduction for both female ($p$ = 0.021) and male speakers ($p$ < 0.001), with overall more reduction in the home space than in the lab space but a larger difference for male (est. −7.99, SE = 1.60) than for female speakers (est. −3.29, SE = 1.37). Figure 5 shows the reduction by gender and space, and the model fit.

## 4. CONCLUSIONS

We found that the vocalization duration not only (trivially) depends on the duration of a session. There was also a significant difference between the vocalization duration in the lab and home space in the informal and the formal experimental condition. The participants in the home group (recorded in Bremen) speak more than the participants in the lab group (recorded in Berlin).

The local speech rate, measured in syllables per second (SPS), differs significantly between informal and formal speech with a slower speech rate in the formal condition, independent of speaker gender or space. This is in line with findings of a slower speech rate in polite and formal speech in other studies [9, 10]. In comparison, fastest SPS-rates obtained from a professional speaker were up to 10 SPS [21]. It also supports the hypothesis that formal situational requirements slow down speech production due to a higher cognitive load while casual and informal conversations facilitate faster speaking rates [11, 12].

For reduction, a significant interaction between speaker gender and space was found, revealing more reduction in the home than in the lab space for both genders but with a larger difference for males. This leads to a gender difference only in the lab condition with females showing more reduction than males. Several acoustic studies have found that male speakers show more vowel reduction and elision, shorter sentence and sound durations, and thus speak faster than females [e.g. 22, 23, 24, 25 and 26, 27, 28]. In our study, females show more reduction than males in the lab setting. One reasoning might be that females are better in adjusting to the situation and fulfilling the communicative tasks, elaborating on a given topic, with only little time to prepare, reflecting this in a more reduction.

All in all, our results point at the fact that situational and functional phonetic variation is not only dependent on the speech task, the speaker (gender) and the addressee, but also on the physical location at which conversations take place. It therefore appears that the role of the specific location and therefore mind-set that the speaker was made believe to be speaking in is a factor worth exploring further. Anecdotal evidence already suggests that there is an effect of place of conversation, as speaking more tamed and quietly in a library or in an elevator is culturally agreed upon in most western societies. In further work we will therefore expand our exploration of the effect of place on phonetic realization.

The reasons for the slower speaking rate in formal situations remain to be analyzed in more detail but potentially comprise processing and memory load constraints connected with formal situational requirements but also factors associated with speaker perception and attributed personality styles such as seriousness or integrity.

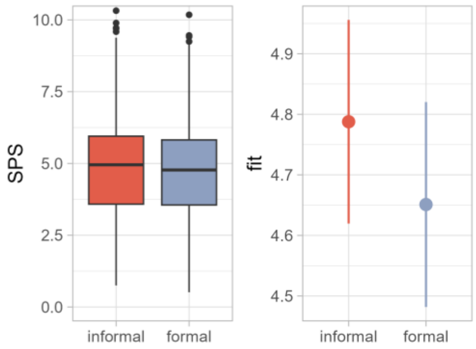

**Figure 4**: Syllables per second by experimental condition (left panel, omitting 13/3889 extreme data points, i.e. outliers, with SPS > 10 from the plot for readability), and the predicted model fit (right panel).

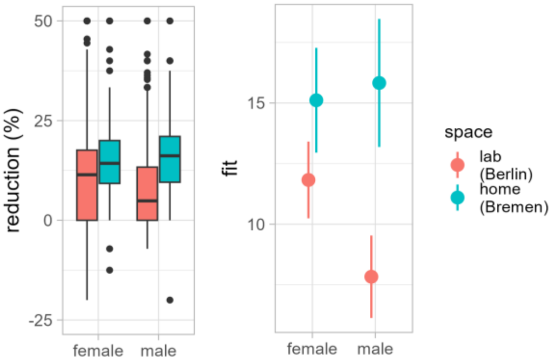

**Figure 5**: Reduction by speaker gender and space (left, omitting 11/3889 extreme data points from the plot with reduction > 50, and 2 with reduction < −25 for readability), and the predicted model fit (right).

## 4. ACKNOWLEDGEMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Council) — SFB 1412, grant 416591334, project C02. Ethics approval was obtained from the Ethics Committee of the Deutsche Gesellschaft für Sprachwissenschaft (DGfS, German Society of Linguistics) under the number #2019-07-190625.

## 5. REFERENCES

[1] A. Bell, “Language style as audience design,” *Language in Society*, vol. 13, no. 2, Art. no. 2, 1984.

[2] J. Hay, S. Jannedy, and N. Mendoza-Denton, “Oprah and /ay/: Lexical frequency, referee design, and style,” in *The Routledge Sociolinguistic Reader*, M. Meyerhoff and E. Schleef, Eds. Routledge, 2010, pp. 53–58.

[3] R. J. Podesva, “Phonation type as a stylistic variable: The use of falsetto in constructing a persona,” *Journal of Sociolinguistics*, vol. 11, no. 4, Art. no. 4, 2007, doi: 10.1111/j.1467-9841.2007.00334.x.

[4] N. Mendoza-Denton, “The Semiotic Hitchhiker’s Guide to Creaky Voice: Circulation and Gendered Hardcore in a Chicana/o Gang Persona,” *Journal of Linguistic Anthropology*, vol. 21, no. 2, pp. 261–280, Dec. 2011, doi: 10.1111/j.1548-1395.2011.01110.x.

[5] S. Jannedy and M. Weirich, “Sound change in an urban setting: Category instability of the palatal fricative in Berlin,” *Laboratory Phonology*, vol. 5, no. 1, Jan. 2014, doi: 10.1515/lp-2014-0005.

[6] W. Labov, The Social Stratification of English in New York City, 2nd ed. Cambridge University Press, 1966/2006. doi: 10.1017/CBO9780511618208.

[7] D. Biber and S. Conrad, *Register, Genre, and Style*. Cambridge, UK: Cambridge, 2009. doi: 10.1017/cbo9780511814358.

[8] V. N. Pescuma et al., “Situating language register across the ages, languages, modalities, and cultural aspects: Evidence from complementary methods,” *Frontiers in Psychology*, vol. 13, 2023, doi: 10.3389/fpsyg.2022.964658.

[9] S. Grawunder, M. Oertel, and C. Schwarze, “Politeness, culture, and speaking task - paralinguistic prosodic behavior of speakers from Austria and Germany,” in *Speech Prosody 2014*, May 2014, pp. 159–163. doi: 10.21437/SpeechProsody.2014-20.

[10] K. Koppen, M. Ernestus, and M. van Mulken, “The influence of social distance on speech behavior: Formality variation in casual speech,” *Corpus Linguistics and Linguistic Theory*, vol. 15, no. 1, pp. 139–165, May 2019, doi: 10.1515/cllt-2016-0056.

[11HUTT] K. H. Huttunen, H. I. Keränen, R. J. Pääkkönen, R. Päivikki Eskelinen-Rönkä, and T. K. Leino, “Effect of cognitive load on articulation rate and formant frequencies during simulator flights,” *The Journal of the Acoustical Society of America*, vol. 129, no. 3, pp. 1580–1593, Mar. 2011, doi: 10.1121/1.3543948.

[12MP] M. K. MacPherson, “Cognitive Load Affects Speech Motor Performance Differently in Older and Younger Adults,” *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 5, pp. 1258–1277, May 2019, doi: 10.1044/2018_JSLHR-S-17-0222.

[13LIV] S. E. Lively, D. B. Pisoni, W. Van Summers, and R. H. Bernacki, “Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences,” *The Journal of the Acoustical Society of America*, vol. 93, no. 5, pp. 2962–2973, May 1993,

[14] T. Kisler, U. Reichel, and F. Schiel, ‘Multilingual processing of speech via web services’, *Computer Speech & Language*, vol. 45, pp. 326–347, Sep. 2017, doi: 10.1016/j.csl.2017.01.005.

[15] P. Boersma and D. Weenink, *Praat: doing phonetics by computer*. 2022. [Software]. Available: http://www.praat.org/

[16] R. Winkelmann, K. Jaensch, S. Cassidy, and J. Harrington, *emuR: Main Package of the EMU Speech Database Management System*. 2021.

[17] S. Kleiner, R. Knöbl, Dudenredaktion (Bibliographisches Institut), and Institut für Deutsche Sprache, Eds., *Duden, das Aussprachewörterbuch, 7. Auflage*. Berlin : Mannheim: Dudenverlag ; Institut für Deutsche Sprache, 2015.

[18] J. Koreman, “Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech,” *The Journal of the Acoustical Society of America*, vol. 119, no. 1, pp. 582–596, Jan. 2006, doi: 10.1121/1.2133436.

[19] 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, doi: 10.18637/jss.v067.i01.

[20] R. V. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means*. 2023. [Software]. Available: https://CRAN.R-project.org/package=emmeans

[21] S. Jannedy, S. Fuchs, and M. Weirich, ‘Articulation beyond the usual: Evaluating the fastest German speaker under laboratory conditions’, in *Between the Regular and the Particular in Speech and Language*, S. Fuchs, P. Hoole, C. Mooshammer, and M. Żygis, Eds., Peter Lang, 2010, pp. 205–234.

[22] D. Byrd, ‘Preliminary results on speaker-dependent variation in the TIMIT database’, *The Journal of the Acoustical Society of America*, vol. 92, no. 1, pp. 593–596, Jul. 1992, doi: 10.1121/1.404271.

[23] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, ‘Acoustic characteristics of American English vowels’, *The Journal of the Acoustical Society of America*, vol. 97, no. 5, p. 3099, 1995, doi: 10.1121/1.411872.

[24] C. Ericsdotter and A. M. Ericsson, “Gender differences in vowel duration in read Swedish: Preliminary results,” *Proceedings of Fonetik 2001, XIVth Swedish phonetics conference. Working papers / Lund University, Department of Linguistics and Phonetics*, vol. 49, pp. 34–37, 2001.

[25] Adrian P. Simpson and Christine Ericsdotter, “Sex-Specific Durational Differences in English and Swedish,” in *15th International Congress of Phonetic Sciences*, 2003, pp. 1113–1116.

[26] S. P. Whiteside, “Temporal-based acoustic-phonetic patterns in read speech: some evidence for speaker sex differences,” *Journal of the International Phonetic Association*, vol. 26, no. 1, pp. 23–40, Jun. 1996, doi: 10.1017/S0025100300005302.

[27] A. P. Simpson, ‘Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung’, *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung (AIPUK)*, 33, 1998.

[28] K. Johnson and J. Martin, “Acoustic Vowel Reduction in Creek: Effects of Distinctive Length and Position in the Word,” *Phonetica*, vol. 58, no. 1–2, pp. 81–102, Jun. 2001, doi: 10.1159/000028489.