Disfluency in Spontaneous Speech (DiSS) Workshop 2023
28-30 August 2023, Bielefeld, Germany

Register differences in ÄH–ÄHM filler particles?

Daniel Duran, Stefanie Jannedy

Leibniz-Centre General Linguistics (ZAS), Berlin, Germany
duran@leibniz-zas.de, jannedy@leibniz-zas.de

Abstract

We present an analysis of register variation (i.e. conventionalized and socially recurring linguistic patterns of intra-individual speech behaviour) of filler particles in German spontaneous speech in contrasting situated interactions. In this study, we analyse intra- and inter-individual differences in the use of lexical type of filler particles and their phonetic realizations. Data was elicited with a novel method where participants talk to a video-taped interlocutor in a simulated tele-conference. This set-up allows us to systematically and consistently manipulate the experimental condition while eliciting laboratory induced differences in fine phonetic detail. We find evidence for register differences but also a high inter-individual variability.

Index Terms: register, phonetic detail, video-interaction task, formality

1. Introduction

Variation in speech style depending on the speaker's addressee and audience has long been observed [1, 2, 3]. Specific studies on cross-situational variation of single speakers (e.g. [4]) have shown that differences in fine phonetic detail may serve to construct differences in social meaning, group affiliation and personas. Such socially recurring situational and functional variation of speech is defined as a speech *register* [5]. Specifically, we refer to differences in speech style due to situational (i.e. level of formality) and functional (i.e. request, narration) variation and perceived social status of the interlocutor as differences in *register* [6].

The study of intra-speaker variation as a field of study has gained traction with the so-called *Third Wave* in sociolinguistics [7], where studies focus on the speech styles of individuals as they maneuver social situations. So far, the study of register has been dominated by work on large corpora of texts [6, 8]. Analyses of register variation in fine phonetic detail are still rare. In our work, we set out to elicit laboratory induced intra-individual stylistic differences in fine phonetic detail. In this study we present an investigation of the production of *filler particles* (FPs, also called *filled pauses*, or *hesitation markers*, cf. [9]) in different registers based on German speech data collected with a novel video interaction paradigm which we have developed [10, 11, 12].

Belz [9] provides a detailed description of filler particles in German, analyzing two different databases [13, 14]. He found that the type of FP and its vowel quality depend on the linguistic context. He also notes that the number of FPs depends on the given situation. However, whether filler particles characterize a *register*, is still an open question [9]. One might assume that a formal speech register induces a higher cognitive load which could lead to more disfluencies [15, 16] — a hypothesis that we test in this paper.

2. Data

We analyze speech recordings collected with a novel experimental paradigm in which we control the addressee persona, the situational set-up and the function of the discourse in a simulated video interaction. Two types of tasks are given to each participant, one of them posing a potential face-threat in which the stakes are high to succeed and successfully negotiate the situation at hand.

The participants saw two pre-recorded videos of a silent female interlocutor who appears in a *formal* scenario and an *informal* scenario with different hair styles, clothing and make-up, sitting at a desk and changing slightly in head, arm and upper body position in a choreographed manner while apparently listening to her interlocutor. The formality of the interlocutor has been assessed in a pre-study. Two stimulus videos were produced based on these previous ratings of perceived personality traits of the depicted interlocutor: one to be perceived as most formal and one to be perceived as least formal [10, 11, 12].

In the formal situation, participants had to either request a deadline extension for a term paper, negotiate a pay raise from a superior or request a rent reduction from the landlady. In the informal situation, participants had to either describe their favourite recipe to a new neighbor, describe their favourite things to do around town to an acquaintance or discuss problems of online courses with a fellow student. These situations and stimulus videos were carefully designed in order to induce the largest possible contrast between the social contexts in terms of formality. The duration of each stimulus video was 2 minutes. We present analyses of 45 participants. All participants gave their written informed consent prior to each recording session and received a compensation of 10 Euros for participation. They identified as either female (25) or male (20).

For this analysis, we distinguish three types of filler particles: vocalic 'V' (as in ⟨ äh ⟩), nasal 'N' (as in ⟨ mm ⟩) and complex 'VN' (i.e. vocalic + nasal, as in ⟨ ähm ⟩). These can occur in four different syntactic/prosodic contexts: 1. within utterances, i.e. between words, (W_W), 2. at the beginning of utterances following a silent pause (#_W), 3. at the end of an utterance preceding a silent pause (W_#), or 4. within a silent pause (#_#). We analyse a total of 1419 filler particles (out of a total of 31424 words; 186.24 min of speech). For the phonetic analysis of vowel quality we take into account 1225 vowels, excluding creaky voice or very short segments.

3. Method

All speech data has been annotated and segmented with WebMAUS [17] and manually corrected in Praat [18]. A Praat script was used to extract the first and second formant F1 and F2 in hertz. All remaining computations have been performed in R [19]. As a measure for vowel dispersion, we computed the center of each speaker's vowel space within the F1×F2 plane and determined the Euclidean distance from the vowel space cen-

10.21437/DiSS.2023-10

ter for each vowel segment. All data analysis and statistical modelling was done in R using the "tidyverse" packages [20]. We fit linear mixed-effects regression models (LMER) with the "lme4" package [21] in combination with "lmerTest" [22]. Post hoc pairwise comparisons were computed using the "emmeans" package [23]. Model comparison statistics have been computed using functions from the "performance" package [24].

In order to compare vowel qualities, we compute *Pillai scores*, a measure which is often used in sociophonetics to assess whether two vowel types are distinct [25, 26, 27]. We compute the Pillai V statistics for all vowel pairs with Multivariate Analysis of Variance (MANOVA) as implemented in the "stats" R library with the first and second formant in hertz as the dependent variable for within-speaker comparisons. Pillai scores range from 0 to 1. Lower values are interpreted in sociophonetics as indicating vowel merger. There is no generally agreed-upon formal criterion for a threshold below which a Pillai score indicates identical or different vowel qualities. We adjust the threshold based on sample size according to the method proposed by [28]. A Pillai score below the sample-specific threshold is interpreted as an indication of vowel merger. Since the threshold is computed separately depending on the compared sets of vowels, we cannot directly compare (or plot) the results across different speakers and sets of vowels. We therefore compute the ratio of Pillai score divided by the adjusted threshold ("p-ratio" in the remainder of the text). The p-ratio is below 1 if the Pillai score is below the threshold and greater than 1 if the Pillai score is above the threshold. We compute Pillai scores only for cases with at least 4 vowel tokens. This leads to some missing values in the vowel quality analysis.

In order to overcome such issues and to obtain a general picture, we fit LMER models for vowel frontness (F2), openness (F1) and dispersion (the Euclidean distance of a vowel token from the speaker's vowel space center). We fit LMER using maximum likelihood (ML) step-by-step by adding fixed effects terms (main effects and interactions) from minimal to maximal. The best fit is determined based on model comparison statistics (ANOVA). As fixed effects, we take into account condition (or speaking style: formal vs. informal), speaker gender and the word-context. As word-context we define three groups: "äh" for all vowels in V-type FPs, "ähm" for all vowels in VN-type FPs, and "W" for all /ɛː/ vowels in stressed positions in lexical words, for reference. As random effects we include SPEAKER.

4. Results

4.1. Frequency of occurrence

Table 1 (a) shows the contingency table for the filler particles by type and formality condition. FP type and formality condition are independent ($\chi^2(2) = 5.08$; $p = 0.079$). VN-type FPs occur more often than V-type FPs, which in turn occur more often than nasal N-type FPs. This is not affected by the formality of the interaction.

Table 1 (b) shows the contingency table for the filler particles by type and speaker gender. We find that the variables filler type and gender are not independent ($\chi^2(2) = 47.541$; $p < 0.001$). Female speakers use more complex VN fillers ("ähm") and male speakers use more vocalic V fillers ("äh"). Nasal fillers are only infrequently used by both gender groups.

Table 1 (c) shows the contingency table for the filler particles by type and context. We find that the variables filler type and context are not independent ($\chi^2(6) = 216.63$; $p < 0.001$).

Table 1: *Number of filler particles by (a) type and condition; (b) type and speaker gender; and (c) type and context.*

(a)

condition	FP type complex	vocalic	nasal	Sum
informal	412	258	35	705
formal	453	221	40	714
Sum	865	479	75	1419

(b)

gender	complex	vocalic	nasal	Sum
female	496	181	35	712
male	369	298	40	707

(c)

context	complex	vocalic	nasal	Sum
W_W	128	211	29	368
#_#	316	46	13	375
#_W	206	142	27	375
W_#	215	80	6	301

4.2. Count LMER

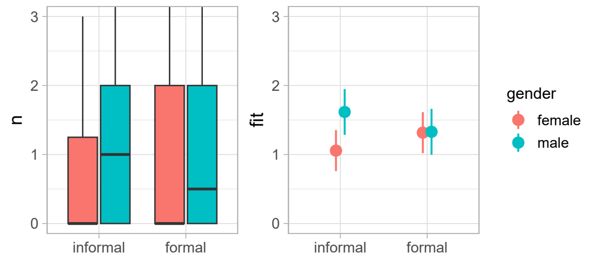

Figure 1: *Count-model. Left: data, right: model fit. Omitted 144 data points with $n > 3$ from plot for readability (out of 1080).*

We fit a LMER model for the number of occurrence (n), with FP-type, speaker gender and condition as fixed effects. The best fitting count-model is:

n ~ TYPE + GENDER + CONDITION + GENDER:TYPE + GENDER:CONDITION + (1| SPEAKER)

As expected from the above statistics, there is an interaction between FP-type and gender. There is also a significant interaction between gender and condition (Figure 1). Estimated marginal means (EMMs) show that the difference between male and female speakers by filler type is significant only for vocalic V-type FPs ($p < 0.001$), but not for VN-type ($p = 0.491$) or nasal N-type FPs ($p = 0.766$). For the difference between male and female speakers by formality condition EMMs are significant only in informal situations ($p = 0.016$) but not in formal situations ($p = 0.956$).

4.3. Vowel quality

Comparisons between the realizations of /eː/ and /ɛː/ in stressed positions within lexical words, i.e. not in FPs, show that there are 20 speakers who merge these two phonemes in

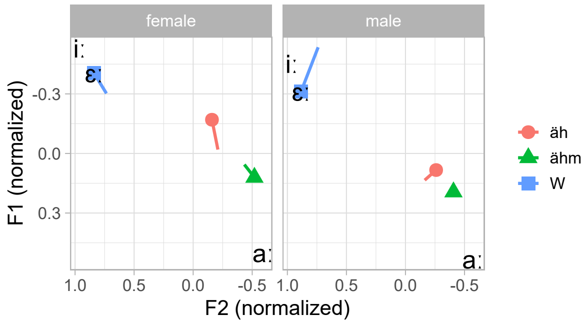

Figure 2: *Vowel quality within the normalized vowel space. Dots and labels (for reference) indicate informal realizations. Line segments point to the realizations in the formal condition.*

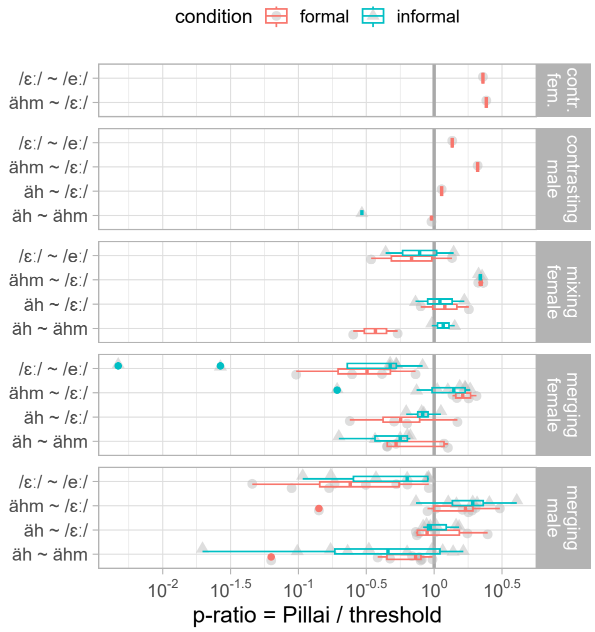

Figure 3: *Vowel comparisons within formality conditions.*

their formal as well as in their informal speaking style (as [eː]). Only 2 speakers contrast these two phonemes in both registers, and another 2 speakers "mix" their realizations. For the remaining speakers there is not enough data. Figure 2 shows the mean locations of stressed /ɛː/ in lexical words ("W"), the V-type ("äh") and the VN-type ("ähm") vowels within the normalized (Lobanov) F1×F2 space (with /iː/ and /aː/ locations for reference). We see that stressed /ɛː/ in words is clearly separated from the vowels in filler particles (which could be expected, since most speakers merge /eː/ and /ɛː/, a common process in Standard German which does not seem to affect filler particles). The differences between V and VN are larger for female speakers. The differences between informal and formal registers (indicated by the line segments), are also larger for female speakers. Both speaker groups have larger register-differences for V-type FPs than for VN-type FPs.

In Figure 3 we see the different vowel comparisons on the vertical axis:
— /ɛː/ ~ /eː/ include all stressed /ɛː/ phonemes in lexical words vs. all stressed /eː/ phonemes in words (for comparison with filler particles);

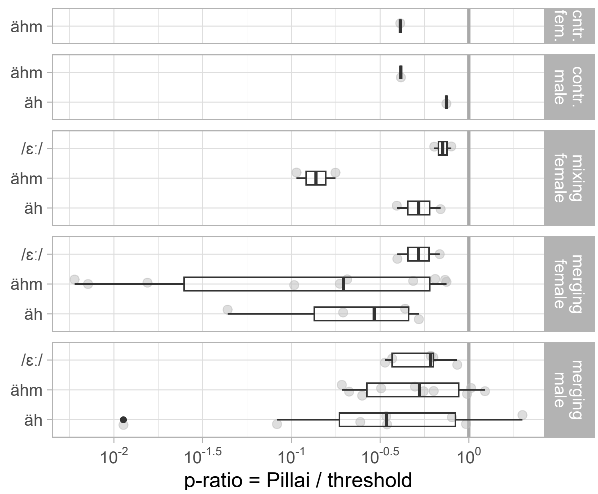

Figure 4: *Vowel comparisons across formality conditions.*

— ähm ~ /ɛː/ include all vowels in VN-type FPs vs. all stressed lexical /ɛː/ segments;
— äh ~ /ɛː/ include all vowels in V-type FPs vs. all stressed lexical /ɛː/ segments;
— äh ~ ähm include all vowels in V-type FPs vs. all vowels in VN-type FPs. /ɛː/ segments.

On the horizontal axis we see the "p-ratio", i.e. the pillai score divided by the threshold. The axis is on a logarithmic scale for better readability. The thick gray line at 10^0 (= 1) indicates Pillai scores exactly at the threshold. Each dot or triangle in the plot corresponds to a subset of vowels by one speaker. The "merging" panels show only speakers who do not distinguish the closed /eː/ from open /ɛː/ in terms of F1 and F2, the "mixing" panels show speakers who distinguish the closed /eː/ from open /ɛː/ in some cases in terms of F1 and F2, and the "contrasting" panels show only the data for speakers who distinguish /eː/ from /ɛː/. As we can see in the rows for the pair "/ɛː/~/eː/", the values in the "merging" panels are all below 1, i.e. these two vowels are merged according to the Pillai scores. Overall, we see indications that the vowels in V-type and VN-type FPs are distinguished by some speakers ("äh ~ ähm"). The vowel in VN-type FPs is distinguished from /ɛː/ by most speakers, while the vowel in V-type FPs is often merged in quality with /ɛː/.

In Figure 4 we see the vowel comparisons across the two speaking styles formal vs. informal (one gray dot per speaker). This is analogous to Figure 3, except that each line along the vertical axis shows a comparison of the same vowel (e.g. the vowel in äh) across speaking styles. We see that most speakers do not distinguish the vowel qualities depending on the formality (p-ratios < 1). Vowels which do not occur in both conditions are omitted from the plot for readability. As the missing items in Figures 3 and 4 suggest, the Pillai scores could not be computed in all cases for all vowel pairs and speakers. Next we model vowel quality for the entire data set using LMER.

4.4. FP-vowel frontness

We model vowel frontness with F2 as the dependent variable, and condition, gender and word-context (W vs V vs VN, WORD) as fixed effects, with random intercepts by speaker. The best fitting frontness-model based on the step-wise forward selection is:

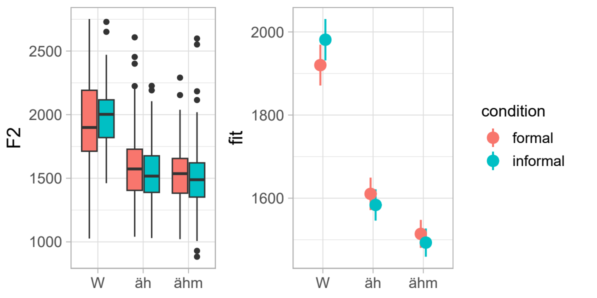

Figure 5: *Frontness-model. Interaction between word-context and formality condition. Left: data; right: model fit.*

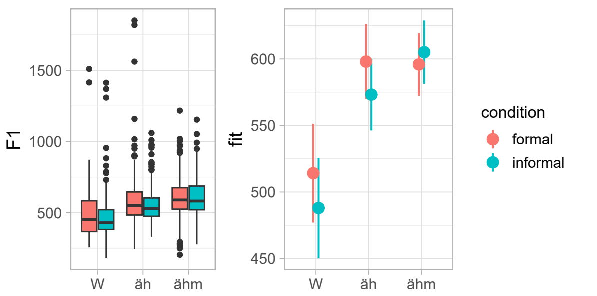

Figure 6: *Openness-model. Interaction between word-context and formality condition. Left: data; right: model fit.*

F2 ~ CONDITION + GENDER + WORD + WORD:GENDER + WORD:CONDITION + (1| SPEAKER)

All main effects are significant, including the main effect for condition (Est $=$ 61.1, SE $=$ 30.5, $p < 0.001$). Additionally, the interaction between condition and word-context is significant ($p < 0.05$, Figure 5). EMMs show that the difference in F2 between the formal and informal conditions by gender and word-context is significant for /ɛː/-vowels in words ($p = 0.036$), but not for V-type FPs ($p = 0.126$) or VN-type FPs ($p = 0.097$). The interaction plot in Figure 5 shows that F2 is higher in ⟨ äh ⟩ and ⟨ ähm ⟩ in the formal condition, although this difference is not significant.

4.5. FP-vowel openness

We model vowel openness with F1 as the dependent variable, and condition, gender and word-context as fixed effects, with random intercepts by speaker.

F1 ~ CONDITION + GENDER + WORD + WORD:GENDER + WORD:CONDITION + (1| SPEAKER)

The main effect for condition is not significant (Est $= -27.1$, SE $=$ 22.25, $p = 0.22$). The interaction between word-context and condition is not statistically significant either (Figure 6). EMMs confirm that there is no statistically significant difference in F1 between the formal and informal condition. The interaction plot in Figure 6 shows reversed patterns for ⟨ äh ⟩ and ⟨ ähm ⟩.

4.6. FP-vowel dispersion

Finally, we fit a model for FP-vowel dispersion (disp). The best fitting model includes the following terms:

disp ~ CONDITION + WORD + WORD*GENDER*CONDITION + (1| SPEAKER)

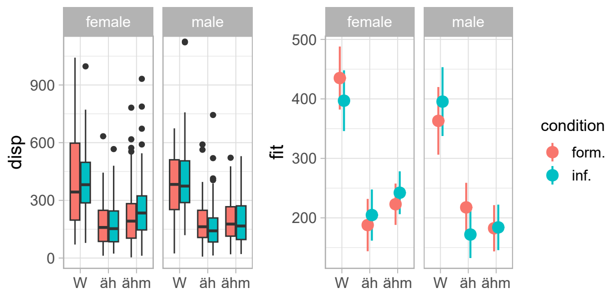

Figure 7: *Dispersion model. Interaction between word-context, gender and formality condition. Left: data; right: model fit. 8 outliers with* disp > 1100 *have been omitted from the data plot for readability.*

The main effect for condition is not statistically significant on its own (Est $= -38.1$, SE $= 30.55$, $p = 0.21$). The three-way interaction between condition, word-context and speaker gender is significant (Figure 7). EMMs show that the difference in dispersion between formal and informal is significant in V-type FPs for male speakers only ($p = 0.007$). The interaction plot in Figure 7 shows reversed patterns for female and male speakers in V-type FPs with female speakers producing a more centralized vowel (i.e. less dispersion) in the formal condition.

5. Discussion

We found that the distribution of filler particles is not only dependent on the context (confirming previous results on German [9]), but that their distribution also depends on the speakers' gender and the speech register (here: formality). Belz [9] found that VN-type FPs are more frequent than V-type FPs for female speakers. We see the same overall pattern of most frequently occurring VN-type FPs, and V-type FPs occurring more frequently than nasal N-type FPs. Notably, female speakers used more "ähm" fillers and male speakers use more "äh" fillers. We found that the FP type (V/VN/N) depends on the syntactic/prosodic context in which it occurs: VN-type FPs occur more often within pauses (#_# contexts) and before silent pauses (W_#), while V-type FPs occur more often between words (W_W) and after silent pauses (#_W). Belz [9] found the same pattern.

Our phonetic analysis of vowel quality showed that the vowels in both VN-type FPs and V-type FPs are distinct from the realization of stressed /ɛː/ in lexical (i.e. non-particle) words. Furthermore, the vowel in V-type FPs is more fronted than the vowel in VN-type FPs. This confirms a (partial) observation reported by Belz. Four out of eight (female) speakers in Belz's study distinguish vowels in VN-type FPs from vowels in V-type FPs in term of vowel quality [9]. We find some evidence for this distinction as well with different patterns depending on register. Overall, there is high inter-speaker variability.

In conclusion, our results confirm previous results for German reported by Belz [9]. Extending these results, we found first evidence for a complex web of interactions between the lexical type of filler particles (V, VN, N), its contexts between words and/or silent pauses, the speaker's gender and speech register (formality). Future studies on filler particles (in German) need to take these factors into account. The new video-interaction paradigm that we developed offers a promising tool for this.

6. Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Council) — SFB 1412, grant-# 416591334, project C02.

We would like to thank two anonymous reviewers for their valuable comments.

7. References

[1] A. Bell, "Language style as audience design," *Language in Society*, vol. 13, no. 2, pp. 145–204, 1984.

[2] D. Burnham, C. Kitamura, and U. Vollmer-Conna, "What's New, Pussycat? On Talking to Babies and Animals," *Science*, vol. 296, no. 5572, pp. 1435–1435, May 2002. [Online]. Available: https://www.science.org/doi/10.1126/science.1069587

[3] J. Hay, S. Jannedy, and N. Mendoza-Denton, "Oprah and /ay/: Lexical frequency, referee design, and style," in *The Routledge sociolinguistics reader*, E. Schleef and M. Meyerhoff, Eds. London; New York: Routledge, 2010, pp. 53–59.

[4] R. J. Podesva, "Phonation type as a stylistic variable: The use of falsetto in constructing a persona," *Journal of Sociolinguistics*, vol. 11, no. 4, pp. 478–504, Sep. 2007.

[5] V. N. Pescuma, D. Serova, J. Lukassek, A. Sauermann, R. Schäfer, A. Adli, F. Bildhauer, M. Egg, K. Hülk, A. Ito, S. Jannedy, V. Kordoni, M. Kuehnast, S. Kutscher, R. Lange, N. Lehmann, M. Liu, B. Lütke, K. Maquate, C. Mooshammer, V. Mortezapour, S. Müller, M. Norde, E. Pankratz, A. G. Patarroyo, A.-M. Pleşca, C. R. Ronderos, S. Rotter, U. Sauerland, G. Schnelle, B. Schulte, G. Schüppenhauer, B. M. Sell, S. Solt, M. Terada, D. Tsiapou, E. Verhoeven, M. Weirich, H. Wiese, K. Zaruba, L. E. Zeige, A. Lüdeling, and P. Knoeferle, "Situating language register across the ages, languages, modalities, and cultural aspects: Evidence from complementary methods," *Frontiers in Psychology*, vol. 13, 2023.

[6] D. Biber and S. Conrad, *Register, Genre, and Style*. Cambridge: Cambridge University Press, 2009.

[7] P. Eckert, "Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation," *Annual Review of Anthropology*, vol. 41, no. 1, pp. 87–100, Oct. 2012.

[8] D. Biber and J. Egbert, "Using Grammatical Features for Automatic Register Identification in an Unrestricted Corpus of Documents from the Open Web," *Journal of Research Design and Statistics in Linguistics and Communication Science*, vol. 2, no. 1, pp. 3–36, Jan. 2016.

[9] M. Belz, *Die Phonetik von äh und ähm: akustische Variation von Füllpartikeln im Deutschen*, ser. Research. Berlin [Heidelberg]: J.B. Metzler, 2021.

[10] E. Brandt, M. Weirich, G. Schüppenhauer, and S. Jannedy, "An experimental approach to cross-situational variation of single speakers," The University of Texas at Austin (virtual), 2021. [Online]. Available: https://www.nwav49.org

[11] D. Duran, M. Weirich, G. Schüppenhauer, and S. Jannedy, "Register Differences in Vowel Dispersion in Formal and Informal Situations," virtual, Apr. 2022. [Online]. Available: https://pretalx.dioe.at/iclave11/talk/ASDSKA/

[12] D. Duran, M. Weirich, and S. Jannedy, "Experimental assessment of intra-individual stylistic variation in fine phonetic detail," submitted for publication.

[13] A. Schweitzer and N. Lewandowski, "Convergence of Articulation Rate in Spontaneous Speech," in *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, Lyon, 2013, pp. 525–529.

[14] M. Belz and C. Mooshammer, "Berlin Dialogue Corpus (BeDiaCo): Version 1," 2020, place: Medien-Repositorium Publisher: Humboldt-Universität zu Berlin. [Online]. Available: https://rs.cms.hu-berlin.de/phon

[15] K. Fischer, "Die Ikonizität der Pause: Zwischen kognitiver Last und kommunikativer Funktion," in *Proceedings der KogWis99*, I. Wachsmuth and B. Jung, Eds. Gesellschaft für Kognitionswissenschaft., 1999, pp. 250–255.

[16] H. Bortfeld, S. D. Leon, J. E. Bloom, M. F. Schober, and S. E. Brennan, "Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender," *Language and Speech*, vol. 44, no. 2, pp. 123–147, Jun. 2001.

[17] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, Sep. 2017.

[18] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," Computer program, 2022, version 6.2.14. [Online]. Available: http://www.praat.org

[19] R Core Team, *R: A Language and Environment for Statistical Computing*, version 4.2.2 (innocent and trusting) ed., R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/

[20] H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani, "Welcome to the tidyverse," *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, 2019.

[21] 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.

[22] 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.

[23] R. V. Lenth, *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2022, r package version 1.8.1-1. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[24] D. Lüdecke, M. Ben-Shachar, I. Patil, P. Waggoner, and D. Makowski, "performance: An R Package for Assessment, Comparison and Testing of Statistical Models," *Journal of Open Source Software*, vol. 6, no. 60, p. 3139, Apr. 2021, r package version 0.10.0.

[25] C. L. Olson, "On choosing a test statistic in multivariate analysis of variance." *Psychological Bulletin*, vol. 83, no. 4, pp. 579–586, 1976.

[26] J. Hay, P. Warren, and K. Drager, "Factors influencing speech perception in the context of a merger-in-progress," *Journal of Phonetics*, vol. 34, no. 4, pp. 458–484, Oct. 2006.

[27] J. Nycz and L. Hall-Lew, "Best practices in measuring vowel merger," Providence, Rhode Island, 2014, p. 060008. [Online]. Available: http://asa.scitation.org/doi/abs/10.1121/1.4894063

[28] J. A. Stanley and B. Sneller, "Sample size matters in calculating Pillai scores," *The Journal of the Acoustical Society of America*, vol. 153, no. 1, pp. 54–67, Jan. 2023.