REDUNDANCY AND INDIVIDUAL VARIABILITY IN THE PROSODIC MARKING OF INFORMATION STATUS IN GERMAN

Janne Lorenzen[1], Simon Roessig[2], Stefan Baumann[1]

[1]I*f*L Phonetik, University of Cologne, [2]Cornell Phonetics Lab, Cornell University
{janne.lorenzen; simon.roessig; stefan.baumann}@uni-koeln.de

ABSTRACT

Information status in German is encoded via prosodic prominence, such that *new* referents are realized prosodically most prominently, *given* referents least prominently and *accessible* referents lie in between these extremes. Since prominence is multifaceted, encompassing a variety of cues related to timing, spectral properties and the F0 contour, one may ask whether speakers are redundant in their prosodic encoding of pragmatic contrasts or select different cues to focus on.

To address this question, we collected data from 32 participants in an interactive reading task with target words varying in information status. We measured six acoustic correlates of prominence related to timing and the F0 contour. Results indicate that some speakers exploit the full range of these cues. However, most speakers reduce this maximally redundant encoding of prominence by following individual strategies. These are compensatory in nature, prioritizing select cues which nevertheless serve the same prominence-marking function.

Keywords: German, individual differences, information status, prominence, prosody

1. INTRODUCTION

Prosodic prominence is multidimensional, encompassing a variety of cues related to timing, changes in F0 and spectral characteristics of the speech signal [1],[2]. Pragmatic categories differentiated via prominence levels are thus often redundantly encoded [3]. To reduce this redundancy, speakers may only select single cues, which gives way to considerable inter-speaker variability in prominence production. In this paper, we investigate how redundant individual speakers are in their prosodic encoding of information status in German.

Inter-speaker variability has recently gathered much attention in prosodic research. For example, [4] show that German speakers differ in the type and number of prosodic cues they employ to distinguish between broad, narrow and contrastive focus. Similarly, [5] conclude that American English (AE)-speaking individuals differ in how strongly they encode informativity (i.e., focus type, contextual probability, and word frequency) via the F0 contour. [6] finds considerable individual differences in boundary marking by AE speakers. Concomitantly, perception studies observe substantial variability in the cues listeners attend to in the decoding of prosodic prominence [1],[4].

Previous studies on individual differences in prominence production have been concerned predominantly with focus marking, a function that in German is typically strongly encoded via prosodic prominence, e.g., the use of different pitch accent types [7]. The present study however looks at information status. Information status in German is often prosodically marked in that *discourse-new* referents are most prominent, *accessible* referents are less prominent and *given* referents least prominent [8],[9],[10]. Since information status can also be signaled via morphosyntactic cues (i.e., *new* referents are preceded by indefinite articles, *accessible* and *given* referents by definite ones), the *given-new* distinction is arguably prosodically less strongly encoded than focus. This allows for even more variability to emerge, as speakers may choose not to encode this contrast via prosodic cues at all.

This paper investigates how differently speakers behave in their prosodic encoding of information status. We will consider several acoustic cues, allowing us to explore the degree of redundancy in prominence encoding, both at population level and individually.

2. METHODOLOGY

2.1. Reading material

To address these research questions, we collected data from a reading task. Target words were disyllabic with stress on the initial syllable. They were embedded in eight different short stories consisting of four sentences each. The third sentence included the target words in indirect and direct object roles. These were either referentially and lexically *new* or *accessible* through the context (see Fig. 1 for an example of a story with an *accessible* indirect object and a *new* direct object). Each story was devised in four different combinations of *new* and *accessible* target words, but participants saw only one version of each story in a Latin square design.

Anna hatte einen erfolgreichen Tag.

"Anna had a successful day."

Am Mittag sind einige Handwerker von der Baustelle nebenan in ihr Geschäft gekommen.

"At noon, some craftsmen from the construction site next door came into her store."

*Unter anderem hat sie **dem Maler eine Waage** verkauft.*

"Among other things, she sold the painter a scale."

Jetzt ist sie zu Hause und entspannt sich bei schöner Musik.

"Now she is at home and relaxes by listening to beautiful music."

Figure 1: Example story with an *accessible* target word ('painter') followed by a *new* one ('scale').

2.2. Experimental procedure and participants

Recordings were collected remotely via a video call with the participant, the experimenter and a confederate. Participants were asked to complete an interactive task with the confederate, in which they read a short story aloud so that the confederate would be able to memorize the story and sort corresponding picture cards into the correct order. During the task, participants wore headphones and recorded themselves onto their own laptops via a podcasting app (Ennuicastr, https://ecastr.com/), which served to circumvent recording issues caused by unstable internet connections. We recorded 32 native speakers of German (8 male, 24 female, 20-38 years old). Each speaker read one version of each story, resulting in 512 target words (256 utterances), which were always produced in broad focus and with a pitch accent.

2.3. Acoustic parameters

We measured three parameters related to the alignment and scaling of the F0 contour (synchrony, tonal onglide, and Delta F0), two parameters related to timing (syllable and word duration) and one integrating both timing and intensity (mass).

Syllable duration and word duration were measured only in (the accented syllable of) non-phrase-final target words, in order to avoid differences due to final lengthening. Periodic energy mass is a measure from the ProPer toolbox [11] which quantifies the area under the periodic energy curve as an integral of duration and intensity. Note that mass in a syllable is considered relative to the average mass of a syllable in the utterance and is therefore unitless. Tonal onglide [12] was measured as the difference in semitones between the tonal target of a pitch accent and a preceding "non-accent tone", as determined by labels annotated for the data following the DIMA guidelines [13]. L*+H accents were excluded from tonal onglide measures, as the most salient F0 movement arguably occurs *after* the tonal target. Synchrony and Delta F0 are measures from the ProPer toolbox as well. In contrast to tonal onglide, these parameters are not dependent on landmark annotations, characterizing the F0 contour without the need for prior labeling. Synchrony is a measure of alignment calculating the distance in milliseconds between the tonal center of gravity [14] and the center of periodic energy mass of a syllable. It captures the extent and direction of the F0 contour as well as its shape. Synchrony is reported relative to the duration of the syllable in percent. Delta F0 characterizes the difference in F0 (in semitones) between two syllables, measured at the center of mass of the accented syllable and the center of mass of the preceding one.

The parameters are either locally confined to the accented syllable (syllable duration, synchrony, periodic energy mass) or characterize the whole target word (word duration, tonal onglide, Delta F0). All parameters were z-scored for further analysis.

2.4. Statistics

We used Bayesian mixed effects linear regression models to query the role each acoustic parameter plays in distinguishing *new* from *accessible* referents. Each model contained information status (levels: *new*, *accessible*) as the predictor and one of the acoustic parameters as the response variable. In addition, we included random intercepts for word and speaker and by-speaker random slopes for information status.

The models were fitted in R [15] using the Stan modeling language [16] via the *brms* package [17]. Four sampling chains ran for 4000 iterations each with a warm-up period of 2000 iterations, yielding a total of 8000 posterior samples per model. We used a weakly informative, normally distributed prior with a mean of zero and a standard deviation of ten for the regression coefficients and default priors supplied by *brms* for the remaining parameters. For each model, we report the estimate β and 90% credible intervals under the posterior distributions as well as the posterior probability that β is larger than zero. We judge there to be compelling evidence for the hypothesis that β>0 if zero is not included in the 90% credible interval and the posterior probability Pr(β>0) is larger than 0.95.

To investigate individual differences in the usage of the different acoustic parameters, we further exploit the random slopes estimates. Random slopes capture the differences in the direction and size of an effect between individuals, which makes them ideal for the analysis of speaker-specific behavior. We ran a hierarchical cluster analysis on the random slopes, following [1] who performed a similar analysis on frequentist regression models. This serves to group together speakers who follow similar strategies in their encoding of information status. Rather than considering each of the 32 individuals separately, which quickly becomes convoluted, the amount of the strategies that need to be considered is thus reduced in an objective manner. Furthermore, we tested for correlations between the random slope values associated with the parameters to find out which cues are used to the same end.

3. RESULTS

3.1. Overall results

Parameter	β	90% CI	Pr(β>0)
Tonal onglide	0.21	[0.08; 0.34]	0.99
Synchrony	0.11	[0.01; 0.21]	0.96
Delta F0	0.35	[0.22; 0.47]	1.00
Word duration	0.12	[-0.02; 0.26]	0.92
Syllable duration	0.06	[-0.07; 0.19]	0.76
Periodic energy mass	0.01	[-0.13; 0.14]	0.53

Table 1: General results of the Bayesian models for each z-scored parameter.

To be able to assess the usefulness of investigating individual differences, we will first consider population-level effects. Table 1 shows the results of Bayesian mixed-effects models for every acoustic parameter. Overall, F0-based cues are more likely to distinguish *new* from *accessible* referents than duration-based cues, with higher Delta F0 constituting the strongest cue to newness. Adhering by the decision criteria determined in Section 2.4, changes in tonal onglide, synchrony, and Delta F0 are reliably predicted by information status, while the other parameters fail to meet these criteria.

3.2. Individual and group behavior

Descriptively, the F0-based parameters are also the ones used most consistently across speakers (see Table 2). 22 speakers mark *new* referents with higher tonal onglide values, 21 speakers produce higher synchrony values in *new* referents and 27 speakers use higher Delta F0 in *new* referents. Only 15 speakers use periodic energy mass, which accords with the model results.

Parameter	# of users	mean acc.	mean new	Δ new – acc.
Tonal onglide (in st)	22	0.66	2.61	1.94
Synchrony (in %)	21	1.73	3.43	1.70
Delta F0 (in st)	27	-0.43	0.58	1.01
Word duration (in ms)	19	384	408	24
Syllable duration (in ms)	17	239	261	22
Periodic energy mass (unitless)	15	1.64	1.90	0.27

Table 2: Number of users of a cue and their descriptive means for *new* and *accessible* referents.

Table 2 also summarizes the average values for *new* and *accessible* target words and the difference between the two for users of a parameter. At least for tonal onglide, Delta F0 and the duration measures, these differences likely have a perceptual reality [18],[19], indicating that they are not just the result of random fluctuations but may be used by the listener to distinguish *new* from *accessible* referents.

To query individual differences and similarities, we ran a cluster analysis on the random slopes of the models reported in 3.1. Following [20], a four-cluster solution is deemed the most appropriate as the one with the highest average silhouette width. However, one cluster consisted of just one speaker so we will report on only three clusters. Two other speakers were excluded from this analysis, one because they produced only phrase-final *new* referents so that the effect of newness on duration could not be estimated and another because the recording was clipped and thus unfit for mass measures.

Figure 2 shows the average random slope values and standard errors per cluster for every parameter. A positive value indicates that this parameter increases from the *accessible* referent to the *new* one for this cluster, negative values indicate that parameter values are lower in *accessible* than in *new* referents.

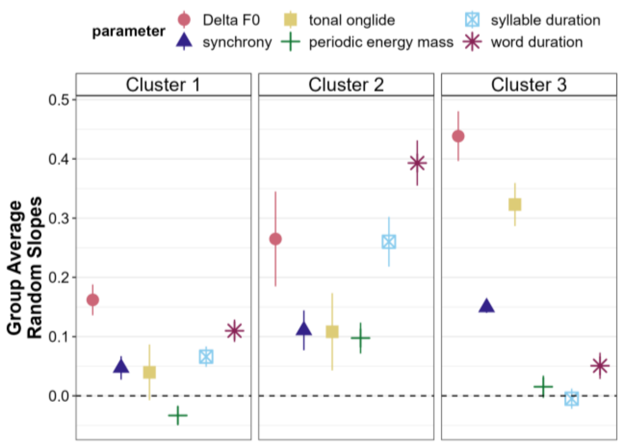

Figure 2: Average random slopes in clusters (speaker groups) with standard errors.

Cluster 1 (11 speakers) includes speakers that produce *new* as opposed to *accessible* referents with higher values in almost all parameters, except for mass. Members of this cluster rely especially on Delta F0, but compared to the other Clusters appear to mark the *new*-*accessible* contrast relatively weakly. Cluster 2 (4 speakers) is defined by a clearer usage of all surveyed parameters to mark *newness*, particularly word duration. Finally, Cluster 3 (14 speakers) relies on F0-based parameters, especially the ones related to scaling, but less on duration and mass. What is also evident from this plot is that no speaker group behaves exactly as predicted by the overall results in Section 3.1. While Cluster 3 somewhat mirrors the general tendency, Clusters 1 and 2 prefer parameters that are not reliably predicted by information status as per the Bayesian models.

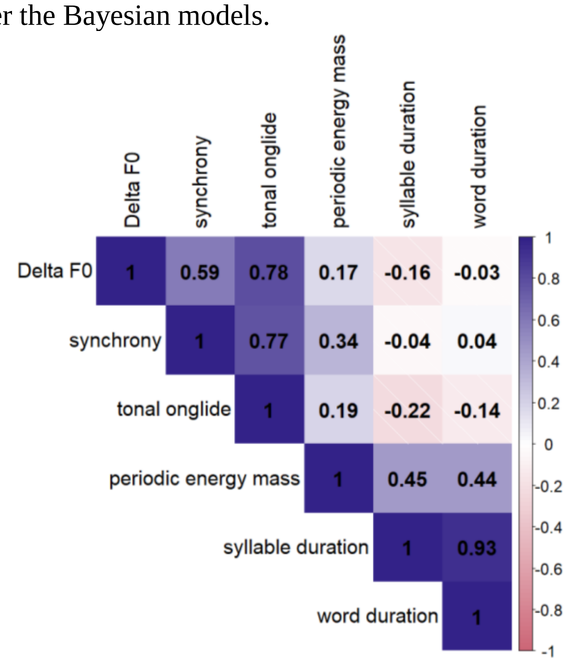

Figure 3: Correlation matrix of random slopes.

We also computed correlations between the random slopes of each model. As evident in Figure 3, tonal onglide correlates fairly strongly with Delta F0 and synchrony, which is unsurprising given that those parameters are meant to capture the shape of the F0 contour. Similarly, syllable duration and word duration are strongly correlated. However, there are no correlations between duration parameters and any of the F0-based cues, indicating that speakers using F0-based correlates tend to not modify duration as a cue to *newness* and vice versa.

4. DISCUSSION

Considering the overall and individual results together, averaging seems to capture the behavior of most individuals quite well, which is reassuring for standard empirical practices. Prosodic cues to prominence, e.g., increased tonal onglide, synchrony and Delta F0, are reliably predicted by *newness* in regression models and are used by the majority of speakers. Other cues, which fail to reach the set reliability thresholds, are concomitantly used by fewer speakers. However, an individual analysis revealed other viable strategies to mark *newness* via prosodic prominence employed by some speakers. For instance, 4 speakers (Cluster 2 in Section 3.2) rely more heavily on longer duration to mark *new* referents as more prominent than *accessible* ones.

Most speakers (18 out of 30, Clusters 2 and 3) mark information status by prioritizing either durational or F0-based cues, which mirrors the finding in [1] that *listeners* tend to focus on either F0-based or durational (plus semantic-syntactic) variables in their processing of prominence. Nevertheless, four speakers (Cluster 2) use all cues investigated here and thus encode information status maximally redundantly and in a cumulative fashion. Thus, redundancy in prominence marking is for the most part compensatory in nature, since it stems from individuals prioritizing different cues serving the same function, but this does not mean that most speakers use a specific prominence cue exclusively.

The robustness of the prosodic encoding of information status has implications for perception: The more redundantly a speaker marks *given* and *new* information in production, the more easily the listener can decode the message [4]. However, it is unclear how the listener deals with contradictory prominence cues, e.g., when a speaker marks *new* referents as more prominent than *accessible* ones via the F0 contour, but at the same time *accessible* referents are longer and thus more prominent in terms of durational cues than *new* ones (or vice versa). A corresponding perception study will reveal the effectiveness of different prominence marking strategies.

5. ACKNOWLEDGEMENTS

We thank Tiffany Zukas for her invaluable help with picture creation and data annotation. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the Collaborative Research Center 1252 “Prominence in Language” (Project-ID 281511265) and project RO 6767/1-1 (Walter Benjamin program).

6. REFERENCES

[1] Baumann, S., Winter, B. 2018. What makes a word prominent? Predicting untrained German listeners’ perceptual judgments. *Journal of Phonetics* 70, 20–38.
[2] Roessig, S., Winter, B., Mücke, D. 2022. Tracing the phonetic space of prosodic focus marking. *Frontiers in Artificial Intelligence* 5.
[3] Winter, B. 2014. Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. *BioEssays* 36, 960–967.
[4] Cangemi, F., Krüger, M., Grice, M. 2015. Listener-specific perception of speaker-specific productions in intonation. In: Fuchs, S., Pape, D., Petrone, C., Perrier, P. (eds.), *Individual differences in speech production and perception*. Peter Lang, 123–145.
[5] Ouyang, I. C., Kaiser, E. 2015. Individual differences in the prosodic encoding of informativity. In: Fuchs, S., Pape, D., Petrone, C., Perrier, P. (eds.), *Individual Differences in Speech Production and Perception*. Peter Lang, 147-189.
[6] Kim, J. 2019. Individual differences in the production of prosodic boundaries in American English. *The Journal of the Acoustical Society of America* 145, 1933.
[7] Baumann, S., Becker, J., Grice, M., Mücke. D. 2007. Tonal and articulatory marking of focus in German. *Proc. 16th ICPhS* Saarbrücken, 1029–1032.
[8] Baumann, S., Grice, M. 2006. The intonation of accessibility. *Journal of Pragmatics* 38, 1636-1657.
[9] Féry, C., Kügler, F. 2008. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36, 680-703.
[10] Baumann, S., Riester, A. 2013. Coreference, lexical givenness and prosody in German. In: Hartmann, J., Radó, J., Winkler, S. (eds.), *Lingua* 136, 16–37.
[11] Albert, A., Cangemi, F., Grice, M. 2018. Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. *Proc. 9th International Conference on Speech Prosody*. Poznań, 804-808.
[12] Ritter, S., Grice, M. 2015. The Role of Tonal Onglides in German Nuclear Pitch Accents. *Language and Speech* 58, 114–128.
[13] Kügler, F., Baumann, S., Röhr, C. T. 2022. Deutsche Intonation, Modellierung und Annotation (DIMA) – Richtlinien zur prosodischen Annotation des Deutschen. In: Schwarze, C., Grawunder, S. (eds.), *Transkription und Annotation gesprochener Sprache und multimodaler Interaktion*. Narr, 23–54.
[14] Barnes, J., Brugos, A., Veilleux, N. Shattuck-Hufnagel, S. 2021. On (and off) ramps in intonational phonology: Rises, falls, and the Tonal Center of Gravity. *Journal of Phonetics* 85, 101020.
[15] R Core Team 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
[16] Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A. 2017. Stan: a probabilistic programming language. *Journal of Statistical Software* 76, 1–32.
[17] Bürkner, P.-C. 2018. Advanced Bayesian multilevel modeling with the R Package brms. *The R Journal* 10, 395–411.
[18] ‘t Hart, J. 1981. Differential sensitivity to pitch distance, particularly in speech. *The Journal of the Acoustical Society of America* 69, 811-821.
[19] Lehiste, I. 1970. *Suprasegmentals*. M.I.T. Press.
[20] Levshina, N. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins.