Does it take two to do an articulatory tango? Investigating the production of novel phonetic forms in varying communicative settings

Leonie Schade* and **Daniel Duran*** and **Sina Zarrieß** and **Joana Cholin** and **Petra Wagner**
Bielefeld University
CRC 1646 – Linguistic Creativity in Communication

Abstract

Successful communication sometimes necessitates creative linguistic solutions. Although speakers are flexible in their choices, they rely on overused planning procedures. Deviating from common pathways requires additional cognitive resources and a good reason for doing so. Current models of word production mainly focus on modelling language behaviour in monologues, falling short of capturing the linguistic innovations that occur in every-day dialogue. In this explorative study, we are interested in speakers' use of novel forms in varying communicative settings, testing the influence of *task*, *setting*, *familiarity*, *syllable frequency* and *personality traits*. Analyzing global phonetic/prosodic features, we find differences between monologues and dialogues within the same speaker on the same task and between different dialogue tasks as well as an effect of individual differences in personality traits. Furthermore, we find signs of increased involvement –or chattiness– in a linguistically easier spot-the-difference game. Lower fundamental frequency ranges in tasks which require more attention to the form, hint at a higher cognitive load. We observe a higher proportion of low-frequency target syllables produced as novel forms and a higher degree of high-frequency syllables produced in canonical patterns. Thus, supporting our expectation of low-frequency syllables to be more susceptible to creative processes than high-frequency syllables.

1 Introduction

Successful communication sometimes necessitates creative linguistic solutions. Although speakers are flexible in their choice of words and structures, they heavily rely on highly overused planning procedures. Deviating from common pathways requires additional cognitive resources and a good reason for doing so, such as attempting to achieve a specific communicative goal. Current cognitive models of word production are mainly focused on modeling highly predictable language behavior in monologue speech, falling short of capturing the linguistic innovations that occur in every-day dialogue. In the current study, we are particularly interested in speakers' use of novel phonetic forms in varying communicative settings.

Research on linguistic creativity at the phonetic level is scarce – there has been some studies on phonetic talent in relation to *language aptitude* and *artistic creative abilities* (e.g. Jilka, 2009). We are, however, not concerned with the exceptional, but rather *everyday creativity* "as an emergent function of dialogue" (Carter, 2015, 13) which is reflected in the production of novel phonetic forms. Previous research shows that spontaneous speech displays a high degree of pronunciation variation (Ernestus and Warner, 2011). Still, phonetic innovations that deviate from the canonical phonotactic inventory of a given language by employing unusual sounds, unusual syllabifications, or unexpected variations, are relatively rare phenomena that are used strategically to aid the communicative goal (Wagner et al., 2021). The investigation of such non-conventional language uses poses a challenge for linguistic theory (Ernestus and Warner, 2011) but also for experimental research: Since creative productions, by definition, do not occur in predictable canonical patterns, they cannot be elicited "directly" from speakers. Instead, novel forms would be expected to occur in spontaneous or task-oriented dialogue settings where speakers are free to deviate from their articulation routines. Yet, detecting such novel forms in spontaneous speech data requires tremendous annotation effort since speech samples need to be transcribed and labeled in a narrow way (which captures fine phonetic detail), whereas common transcription procedures provide orthographic transcriptions, reflecting canonical/citation forms. However, the *dual*

*LS and DD contributed equally to the paper.

Proceedings of the 29th Workshop on the Semantics and Pragmatics of Dialogue, September 3–5, 2025, Bielefeld, Germany.

route account of phonetic encoding (Levelt et al., 1999; Walsh et al., 2010; Cholin et al., 2011) provides a framework explaining how speakers can deviate from more routinized pathways and produce novel phonetic forms. Within this account, speech may either be produced by accessing precompiled motor programmes from a repository (the mental syllabary) via a “retrieval route”, or it may be assembled from smaller units via an “assembly route”. While the retrieval route is assumed to be the faster, more automatized and more efficient pathway, more often used in the production of highly trained, high-frequency syllables, the assembly route is a more controlled pathway which requires more resources but offers a higher flexibility, allowing for the construction of less trained, rare syllables, and even novel pronunciations. Given that rarely occurring syllables are more prone to be generated via the assembly route, it is possible that they are also more susceptible for being the target of innovative modifications. Within the dual route framework, we expect novel pronunciation patterns to involve a higher degree of cognitive load.

In our study, we investigate whether this notion also extends to creative pronunciations, and whether dialogue-based interactions (as opposed to monologues), potentially increase speakers’ tendency for employing novel or unusual pronunciation strategies. This idea receives further support from the finding that speaking in a dialogue differs from speaking in a monologue (Kuhlen and Rahman, 2017), and may in some sense even be easier due to interactive priming effects (Garrod and Pickering, 2007, 2013). Speaking in dialogues may foster the creative potential of individuals when certain aspects of interpersonal dynamics are met: group diversity, social and cognitive stimulation (Paulus, 2000). We, therefore, ask whether varying communicative settings affect the production of novel phonetic pronunciations. In particular, we investigate the following research questions:

1. Do we find evidence for more or less cognitive load (e.g., pauses or hesitations) across varying communicative settings and tasks, which is likely to correlate with creative processes?

2. Do we find evidence for interpersonal dynamics that may foster or inhibit creative pronunciation behaviors (e.g., “chattiness", interpersonal alignment, emotional involvement, personality features)?

3. Do we find evidence for more or less novel pronunciations across varying communicative settings or tasks (e.g., a higher or lower frequency of non-canonical, or novel productions)?

4. Do we find an effect of syllable frequency of occurrence on the number of novel pronunciations (i.e. are low-frequency syllables more affected by novel pronunciation)?

We investigate these research questions by analyzing spontaneous speech productions in German across different tasks and settings: the *Diapix* task (DPX) (Baker and Hazan, 2011), a *password obfuscation task* (PWO), a *product naming* task, carried out both in monologues (PNM) and dialogues (PND), and a *debriefing and interview phase* (DBI) (see Section 2).

Research questions 1 and 2 will be addressed in Section 3 by analyzing global phonetic/prosodic parameters of interaction, indicating the level of *cognitive load*, but also the level of *involvement* across these three tasks, assuming that creative processing will show more signs of cognitive load, and that dialogues show more involvement than monologues. Cognitive load is positively correlated with the frequency of occurrence of speech pauses, hesitations and with longer (filled or silent) pauses (Betz et al., 2023). Involvement will be investigated by looking at turn-internal pauses as well as pitch range (Wrede and Shriberg, 2003; Wagner et al., 2024). As creative involvement may also be driven by interpersonal dynamics and personality related factors, we also assess the influence of *speaker familiarity*, and Big Five personality traits, concentrating our present analysis on *openness to experience* (Jirásek and Sudzina, 2020), as it is the most robust trait related to creative achievement (Ahmed and Feist, 2021).

Research questions 3 and 4 will be addressed in Section 4. The third research question will be investigated by comparing the number of phonetic innovations across tasks. We expect a larger amount of novel forms in the product naming tasks than in the Diapix task based on participants’ feedback reported in Duran et al. (2025), saying that “they understood the DPX [...] as requiring ‘precise’ use of language in contrast to the other tasks, [...] requiring ‘creative’ use of language” (p. 90). We also expect interactive communicative settings to lead to more variation and, thus, innovation (i.e.

more novel productions in dialogues than in monologues). The fourth research question will be investigated by analyzing the interplay of syllable frequency and the probability of its being realized in a canonical or novel fashion. We categorise syllable realisations as novel if they show unexpected variations, re-syllabifications, phones that are not part of the language's inventory, or phonotactic innovations similar to the characteristics described in Wagner et al. (2021) and expect to find a higher degree of target syllables with a low frequency of occurrence to be produced in a novel way.

2 Experimental Design

The experimental methodology of the data used in this study is presented in detail in Duran et al. (2025). As novel phonetic forms, as defined above, are/can be a rare phenomenon, we designed a battery of tasks specifically to encourage the elicitation of novel phonetic forms. The evaluation of the tasks' suitability, i.e. if and how many novel forms were produced is addressed in Section 4. The item set consists of German syllables with either a high- or low-frequency of occurrence (based on corpus data compiled by Samlowski, 2016). The final item set contains 47 high- and 33 low-frequency target items (syllables), including 15 high- and 15 low-frequency syllables from syllabic quartets, following the construction procedures of Cholin et al. (2011). Additionally, socio-demographic meta data (age, gender, language backgrounds etc.) of each participant was collected and personality traits assessed using the Big Five inventory (BFI-10, Rammstedt et al., 2014). 23 participants (13 female, 1 non-binary/diverse, 9 male), between 18 and 32 (mean = 24, median = 21) years, all native speakers of German, participated in 12 dyadic sessions. Seven dyads were peers (friends or acquaintances) and five dyads consisted of strangers. The following tasks were employed with differing task orders:

Diapix (DPX). The Diapix task is an elicitation method for (quasi-) spontaneous, interactive speech in which two participants verbally engage in a spot-the-difference game (Baker and Hazan, 2011). Our target items are incorporated in the depictions on the images such that participants are encouraged to produce these syllables without being told to do so explicitly.

Product Naming Dialogue (PND). Here, two participants were tasked with finding a name for a fictitious product using the two syllables provided orthographically as a starting point. They received 60 products to name in random order. The nature of the task asks participants to "play" with the provided syllables, encouraging novel creations.

Product Naming Monologue (PNM). In the monologue version of the product naming task, participants followed the same instructions as in PND. They were instructed to think aloud while coming up with a name on their own.

Password Obfuscation Task (PWO). This task is another gamification scenario. It involves two participants who have to verbally communicate a password / pass-phrase to their interlocutor in a simulated "man-in-the-middle attack". As they have to find strategies to hide the passwords / pass-phrases (containing the target items) in a way the third person cannot understand, we expect novel strategies and novel productions.

Debriefing & Interview (DBI). After all tasks have been completed, we conducted a short verbal interview with the participants along with the final debriefing. The two participants were seated in the recording lab and the experimenter joined them to talk about their experience with the various tasks.

2.1 Data preparation and analysis

Data annotation and acoustic analyses were done with Praat (Boersma and Weenink, 2025), incorporating automatic transcriptions with BAS web services (Kisler et al., 2017). The produced target syllables are currently being annotated manually by identifying the original target syllable and their *production type*, i.e. if they were produced in their canonical form or in a novel way. Following the findings of Wagner et al. (2021), we consider productions as novel when they show (1) unexpected variations, (2) novel re-syllabificatons of lexemes and (3) phonotactic or allophonic innovations.

We model all global effects statistically in R with linear mixed-effects regression (LMER) and the novel phonetic forms with generalized linear mixed-effects models (GLMER) using *lme4* (Bates et al., 2015) together with *lmerTest* (Kuznetsova et al., 2017), *emmeans* (Lenth, 2022) for post-hoc computations of estimated marginal means (EMMs, i.e. adjusted predictions) for pairwise comparisons of categorical variables.

In LMER models, we consider the speaker ID and the recording session as random effects. As categorical fixed effects we consider the following (the first mentioned category is defined as the base level at the models' intercept): *task* (PNM, DPX, PND, PWO, DBI); *familiarity* (strangers vs. peers); *PN* (first vs. second, encoding whether the participants did the monologue product naming task before the dialogue product naming). For some models, we also take into account *final* (int vs. fin, encoding whether an IPU is turn-internal or turn-final). As numerical fixed effects, we consider the Big Five subscales *extraversion*, *agreeableness*, *conscientiousness*, *neuroticism*, *openness* individually, but only report results for *openness*.

To find the best-fitting LMER models (estimated using REML and nloptwrap optimizer), we apply step-wise addition of variables, starting bottom-up with an intercept only model and then step-by-step adding fixed main effects and interactions until the model fit cannot be improved. We use *influence.ME* (Nieuwenhuis et al., 2012) to remove individual overly influential observations from the model data. For all significance tests, we apply $\alpha = 0.05$.

All GLMER models (estimated using ML and Nelder-Mead optimizer), were fitted so the *production type* (canonical vs. novel) is predicted with the *speaker ID* and the *target syllables* as random effects and either the communicative *setting* (monologue vs. dialogue) or the *task* (DPX, PND, PNM) and the *target syllable frequency* (high vs. low) as fixed effects. Standardized parameters were obtained by fitting the model on a standardized version of the dataset. 95% Confidence Intervals (CIs) and p-values were computed using a Wald z-distribution approximation.

3 Global Effects

	obs.	mean	sd
pause.dur	7511	1.04	0.99 (seconds)
n.IPU	5680	1.71	1.14 (count)
f0.range.z	14849	1.26	1.17 (z-score)
endf0.rg.z	5569	0.87	1.03 (z-score)

Table 1: Descriptive statistics for the dependent variables: the total number of observations (obs.; i.e. the number of data points submitted to the statistical analysis), the mean and standard deviation (sd). Note: The total number of observations differs due to model-dependent removal of overly influential observations.

We first analyze acoustic-phonetic features related to timing and coordination like pause duration or the number of IPUs, resp. pauses per turn (Tab. 1). These provide clues to potentially increased cognitive load, but also a higher degree of a speaker's involvement, due to underlying *creative* speech production processes.

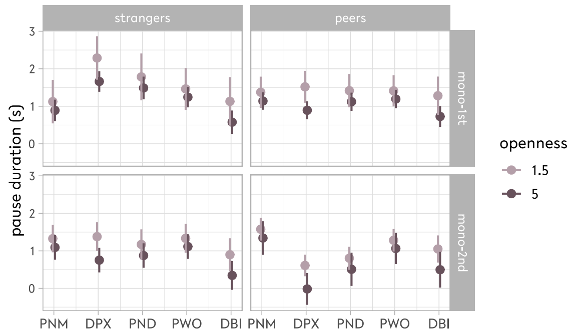

Figure 1: Predicted values of pause duration (by *task*, *familiarity*, product naming task order and *openness* [at min and max values]).

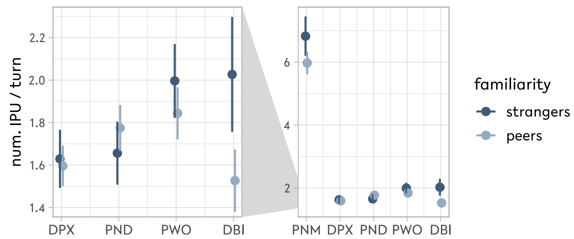

Figure 2: Predicted values of IPUs/turn. The left panel zooms into the smaller differences between the interactive settings.

In the absence of a linguistic analysis, we segment the recorded discourse of each task into *inter-pausal units* (IPUs, i.e. stretches of speech which are separated by a pause). Consecutive IPUs are grouped into *"turns"* if they are not separated by a pause longer than 5 seconds or a speaker change, excluding single short IPUs (cf. Heldner et al., 2011) from the interlocutor.

3.1 Pause durations

The number and duration of pauses may be seen as a potential sign of increased cognitive load. We first analyze the duration (*pause.dur*). As *pauses* we define all turn internal stretches of silence no longer than 5 seconds (at which threshold we assume the start of a new turn). The analyzes are based on the manually checked annotations of IPUs.

Results: Figure 1 visualizes the model predictions. The explanatory power of the best-fitting model is moderate (conditional $R^2 = 0.13$, marginal $R^2 = 0.08$; see Table 4 in the appendix for full details). The EMM results (Fig. 11) for pairwise comparisons involving PNM and DPX are almost complementary for strangers who did the monologue task first, on the one hand, and peers who did the dialogue task first, on the other: (1)

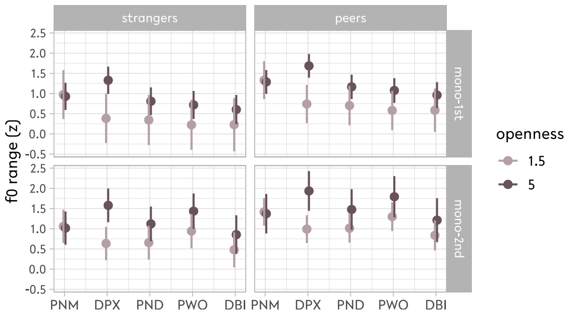

Figure 3: Predicted values of f0 range by *task*, *familiarity*, *PN* task order (monologue first vs. monologue second), and *openness* [at min and max values].

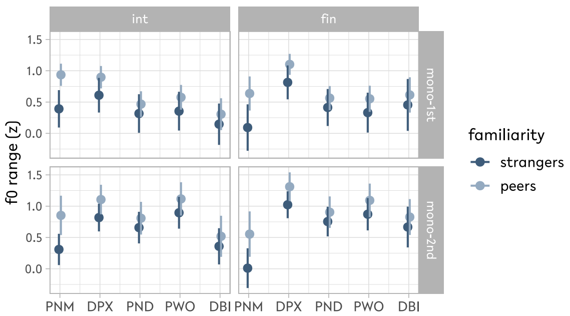

Figure 4: Predicted values of f0 range at IPU ends. Left panels: turn internal IPUs; right: turn-final.

PNM has significantly shorter pauses than DPX, PND, PWO for strangers in the *PN*=first condition. (2) PNM has significantly longer pauses than DPX, PND, DBI for peers in *PN*=second condition. The Diapix task shows a similar, but somewhat reversed pattern: (1) DPX has significantly longer pauses than PWO and DBI for strangers in *PN*=first, and (2) DPX has significantly shorter pauses than PND, PWO and DBI for peers in *PN*=second. Four of the Big Five subscales are included in the model, interacting with *task*, including *openness* for which we see generally shorter pause durations predicted with higher values of openness for all tasks.

3.2 IPUs per turn

After analyzing the duration of pauses, we now turn to their number, as another proxy for cognitive load. Here we analyze how many IPUs there are per turn (variable *n.IPU*).

Results. Figure 2 visualizes the model predictions. The explanatory power of the best fitting model is moderate (conditional $R^2 = 0.20$, marginal $R^2 = 0.18$; see Tab. 5). EMMs (Fig. 12) show that PNM has significantly more IPUs/turn than any other (interactive) task, independent of familiarity. Within the dialogue tasks, we find an effect of familiarity: (1) For strangers: DPX has significantly less IPUs/turn than PWO or DBI; and PND has significantly less IPUs/turn than PWO or DBI. (2) For peers: DPX has significantly less IPUs/turn than PND or PWO; and PND has significantly more IPUs/turn than DBI.

3.3 f0 range

As a proxy for *creative* cognitive speech production processes, potentially indicated by a higher degree of involvement, we analyze variations in fundamental frequency (f0). We are not interested in absolute inter-speaker differences, but in intra-speaker dynamics and variability across the different interactional situations. Thus, in order to be able to compare f0 variations across speakers, we normalize f0 values from the original Hertz scale to z-scaled values by each speaker individually. We model normalized f0 range by computing the interquantile range from 5% to 95% for each interval.

Results. Figure 3 visualizes the model predictions. The explanatory power of the best fitting model is weak (conditional $R^2 = 0.13$, marginal $R^2 = 0.07$; see Tab. 6). Pairwise EMM comparisons (Fig. 13) show two different patterns depending on the product naming task order: (1) for *PN*=first, the f0 range is significantly larger in PNM in comparison to PND, PWO and DBI. (2) for *PN*=second, the f0 range is significantly smaller in PNM in comparison to DPX and PWO. For *PN*=first, speakers also produced a larger f0 range in DPX in comparison to PND, PWO and DBI. For *PN*=second, DPX has also a significantly larger f0 range than PND.

3.4 IPU-end f0 range

We model normalized f0 range at the end of IPUs (*endf0.rg.z*). To do this we extract the final 500ms from each IPU which is longer than one second. We also take into account positional effects, and encode whether an IPU occurs at the end of a turn (variable *final*).

Results. Figure 4 visualizes the model predictions. The explanatory power is moderate (conditional $R^2 = 0.15$, marginal $R^2 = 0.11$; see Tab. 7). EMMs (Fig. 14 show that the PNM task has a significantly smaller f0 range in comparison to all dialogue tasks in turn-final IPUs for stranger in the PN=second condition. This general tendency towards smaller f0 range in PNM is also true for turn-internal IPUs, but the differences are statistically significant only for PNM vs. DPX and PNM vs. PWO.

3.5 Preliminary summary: global features

We find clear differences between the monologue and dialogue settings within the same speaker on the same task (product naming). Furthermore, the acoustic-phonetic global features of the monologue task (PNM) are different from all dialogue tasks. PNM has overall more pauses and a smaller f0 range at the end of IPUs than the dialogue tasks. We also find an effect of familiarity: The pause duration in the Diapix task (DPX) is longer for strangers in comparison to the other dialogues and shorter for peers. We find an effect of task order in the product naming tasks (mono first vs. second). We find that individual differences in personality traits affect the analyzed features. Finally, all variables included in the LMER models interact with *task* — i.e. the models which include an interaction with task always resulted in a better model fit in comparison to models which have only the corresponding main effect.

4 Novel Phonetic Forms

To assess the elicitation suitability of the tasks, we look at the *production type* (canonical or novel) of the uttered target syllables and their distribution across tasks. The differences in production type in monologues vs. dialogues are analyzed on the entire subset and also separately for both product naming tasks. To investigate whether syllables produced via the assembly route are more likely to be subject of creative innovation, we analyze the uttered target syllables' frequencies (high vs. low) in regard to their relation to the production types. For the following analyses, we use a subset of the data that where the production type has thus far been annotated. It consists of the recordings of seven participants (1 d, 2 f, 4 m) in three of the piloted tasks: PNM, PND, DPX (Tab. 2). They produced 1224 instances of 51 different target syllables in total, six of which were excluded because of unintelligibility and signal distortion, yielding a set of 1218 target syllables; 514 of these are canonical and 704 novel productions (42.20% and 57.80%, respectively). One participant, P173, blended/merged together target syllables in PNM. For the present analyses, these are treated as their separate target syllables. Table 2 shows the number of uttered target syllables and their production type, canonical or novel, for each participant and across the three tasks (Fig. 15). The amount of produced target syllables in each task varies from 11.90% in DPX to 59.85% in PNM, therefore, the distribution of canonical and novel productions is computed within each of the tasks.

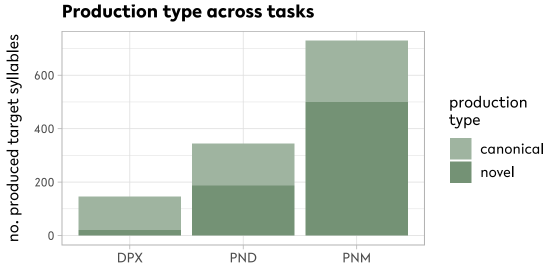

Figure 5: Number of uttered canonical and novel productions across the three tasks DPX, PND and PNM.

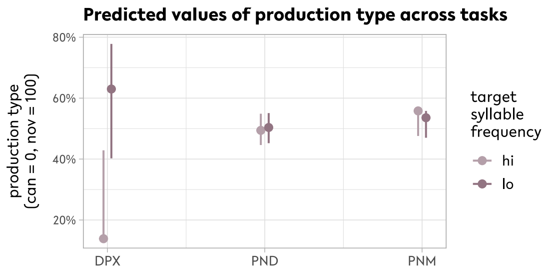

Figure 6: Predicted values of canonical and novel productions of high ("hi") and low ("lo") frequency syllables within tasks with the production type canonical = 0 and novel = 100.

4.1 Tasks

Looking at the three tasks individually (Tab. 2, Fig. 5), PNM has the highest percentage of novel productions (68.45% of productions within the task, n = 499), followed by PND (54.07%, n = 186) with DPX showing the lowest percentage of novel productions (13.10%, n = 19).

The GLMER to predict *production type* with the *task* and *target syllable frequency* (Tab. 12) has a substantial total explanatory power (conditional $R^2 = 0.38$) and the part related to the fixed effects alone (marginal R^2) is of 0.22. The model's intercept, corresponding to task[DPX] and target_syl_freq[hi], is at -2.78 (95% CI $[-3.68, -1.88]$, $p < 0.001$). All variables and interactions have a statistically significant effect of $p < 0.001$. The predicted values are visualized in Figure 6.

task & prod. type	P164	P173	P252	P317	P425	P517	P724	total	ratio prod. type within tasks (%)	distribution of prd. syls (%)
DPX	16	35	21	15	25	18	15	145		(11.90)
can	14	28	16	15	21	17	15	126	86.90	24.51
nov	2	7	5	0	4	1	0	19	13.10	2.70
PND	9	33	33	68	120	25	56	344		(28.24)
can	5	18	9	67	30	14	15	158	45.93	30.74
nov	4	15	24	1	90	11	41	186	54.07	26.42
PNM	35	68	188	112	162	37	127	729		(59.85)
can	19	32	52	58	35	9	25	230	31.55	44.75
nov	16	36	136	54	127	28	102	499	68.45	70.88
total	60	256	122	195	307	80	198	1218		
can	38	98	57	140	86	40	55	514	*42.20*	
nov	22	158	65	55	221	40	143	704	*57.80*	

Table 2: Distribution of novel and canonical productions of target syllables per participant within the different tasks. The percentages in italics refer to the ratio of total canonical and novel productions among all productions. The rightmost column shows the distribution of the produced syllables (prd. syls) among the tasks. The values in brackets refer to the ratio of a task's total amount of productions among all productions.

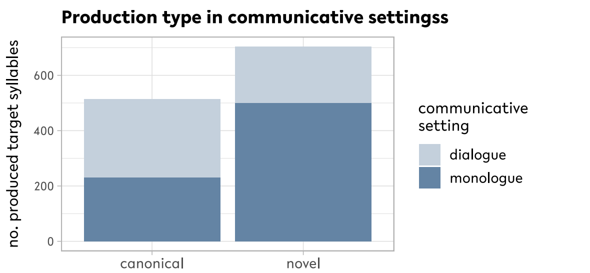

Figure 7: Number of uttered canonical and novel productions within the dialogue (DPX, PND) and monologue (PNM) tasks.

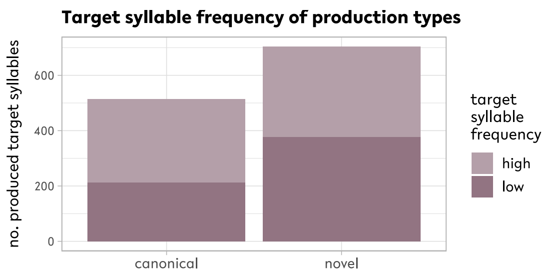

Figure 9: Number of uttered high- and low-frequency syllables within canonical and novel productions.

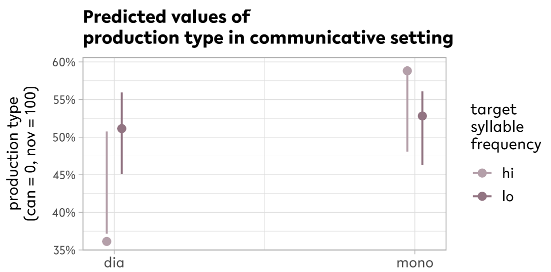

Figure 8: Predicted values of canonical and novel productions of high- ("hi") and low -("lo") frequency syllables in mono- and dialogues with the production type canonical = 0 and novel = 100..

4.2 Varying communicative settings: mono- and dialogues

Figures 7 and 16 and Table 8 show the dataset with both dialogue tasks merged into one category and the PNM task as the monologue category. The percentage of novel productions within monologues remains 68.45% (n = 499), while 41.92% (n = 205) were produced in the dialogue tasks.

The GLMER to predict the *production type* with the communicative *setting* and *target syllable frequency* (Table 11) has substantial explanatory power (conditional $R^2 = 0.37$) and the part related to the fixed effects alone (marginal R^2) is 0.16. The model's intercept, corresponding to monodia[dia] and target_syl_freq[hi], is at -1.37 (95% CI $[-2.12, -0.61]$, $p < 0.001$). The effects of all variables and interactions are statistically significant ($p < 0.001$). The predicted values are plotted in Figure 8.

A model comparison of the GLMER using *task* vs. *setting* as a fixed effect with ANOVA reveals the latter to have significantly lower AIC and BIC values and, thus, have a better model fit (Tab. 13).

4.3 Syllable frequency

53.41% of novel productions originated from a target syllable with a low frequency of occurrence (n = 376), while 58.56% of canonical productions were high-frequency syllables (n = 301) (see Tab. 3, Fig. 9 and 17). Overall, 51.64% of uttered target syllables had a high frequency (n = 629) and 48.36% had a low frequency (n = 629).

prod. type & syl freq	P164	P173	P252	P317	P425	P517	P724	total	ratio syl freq in prod. type (%)
canonical	38	78	77	140	86	40	55	514	(42.20)
high	27	56	39	67	55	25	32	301	58.56
low	11	22	38	73	31	15	23	213	41.44
novel	22	58	165	55	221	40	143	704	(57.80)
high	15	16	84	35	102	20	56	328	46.59
low	7	42	81	20	119	20	87	376	53.41
total	60	136	242	195	307	80	198	1218	
high	42	72	123	102	157	45	88	629	*51.64*
low	18	64	119	93	150	35	110	589	*48.36*

Table 3: Distribution of high- and low-frequency target syllables per participant across novel and canonical productions. Percentages in italics refer to the ratio of total high- and low-frequency productions among all productions. Values in brackets refer to the ratio of the production type among all utterances.

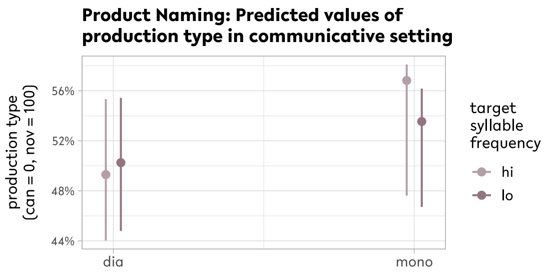

Figure 10: Predicted values of the production type of high- ("hi") and low- ("lo") frequency syllables in the mono- and dialogue version of the product naming task with the production type canonical = 0 and novel = 100. Only the effect of monodia[mono] is statistically significant (Tab. 14).

4.4 Product naming subset

Within the subset of the product naming tasks, 63.84% of uttered target syllables were produced as novel phonetic forms and 53.87% of uttered target syllables had a low frequency (Tab. 9 and 10, Fig. 18). A GLMER was fitted to predict the *production type* with the communicative *setting* and *target syllable frequency* (Tab. 14). The model's total explanatory power is moderate (conditional $R^2 = 0.25$) and the part related to the fixed effects alone (marginal R^2) is of 0.03. The model's intercept, corresponding to monodia[dia] and target_syl_freq[hi], is at -0.15 (95% CI $[-0.96, 0.65]$, $p = 0.708$). Only the effect of monodia[mono] is statistically significant ($p = 0.004$). The predicted values are plotted in Figure 10.

4.5 Preliminary summary: novel phonetic forms

We find most novel forms to have been produced in the PNM task, fewest in DPX and that the monologue task led to more novel elicitations of novel forms than both dialogue tasks combined. The variables *task* or *setting* and *target syllable frequency* are significant in predicting the *production type* with the model using *setting* having a better model fit. Furthermore, low-frequency syllables are more likely to be produced with novel pronunciations, while high-frequency syllables are more likely to be produced canonically. Focusing on the product naming tasks, 63.84% of productions here are novel and 53.87% of productions have a low frequency.

5 Discussion & Conclusion

In this paper, we focused particularly on the question of whether the production of novel phonetic forms varies depending on the communicative setting, i.e. monologue vs. dialogue tasks.

We find not only clear differences between monologues and dialogues within the same speaker during the same task (product naming) but also within the same speaker across the different dialogue tasks. The global phonetic-linguistic features of the monologue product naming task are clearly different from the other dialogue tasks. The results from the "global" analyses (Section 3) show that the duration of pauses depends on the familiarity of the speakers and the tasks — with differences between mono- vs. dialogue as well as the different interactive tasks. The Diapix task (DPX) has longer pauses than the other dialogue tasks for strangers and shorter pauses for peers (friends or acquaintances). The monologue product naming task (PNM) has more pauses than the dialogue tasks. Within the dialogues, we found that DPX has less pauses per turn than the other tasks. This could be interpreted as an indicator of increased involvement or chattiness — with more turn-taking in the linguistically easier spot-the-difference game. In addition, familiarization with the task affects speech production, as evidenced by the effect of task order in product naming (mono first vs. second) on the range of the fundamental frequency (f0). The f0 range at the end of IPUs is smaller in the monologue task in comparison to the other tasks, highlighting the communicative function of intonation in interactional settings, and potentially indicting a higher cognitive load.

In addition, we find that individual differences in personality traits (as captured by the Big Five inventory scales) affect the analyzed features. Interestingly, all variables included in the LMER models interact with *task* — i.e. the models including an interaction with task always resulted in a better model fit in comparison to models which included only the corresponding main effect, further highlighting the that the communicative setting affects speech production.

The tasks developed by Duran et al. (2025) are suitable to elicit novel forms, as the results clearly show: 57.80% of all uttered target syllables were novel productions. When we look only at the 'creative' product naming tasks, the amount is even higher: 63.84% of all productions are novel (Tab. 9). The DPX results lead us to interpret it as a suitable task for the production of spontaneous dialogue and will in future analyses rather serve as a baseline to compare the other tasks to.

The analyses on the novel phonetic forms (Section 4) show that more novel forms were produced in the product naming tasks than in DPX and that PNM elicited the most novel productions, i.e. the tasks encouraging to be creative on the linguistic/phonetic level result in more novel productions than the spot-the-difference task. This corroborates our expectation that DPX is different from the other tasks and supports the results found in Section 3 as well as the qualitative results from Duran et al. (2025).

Contrary to our initial expectation, the highest percentage of novel forms was produced in the monologue task (PNM). In terms of the product naming task, however, this seems plausible. We believe some confounding factors could have led to this high percentage in PNM: on the one hand, the participants might have been less inhibited to produce weird or nonsense forms and could have just mumbled things because there was no listener present and they were, as instructed, merely "thinking aloud". Therefore, there was no need to produce something that is up to the standard and underlies the criteria for listener oriented speech. Rather, speakers could already throw works-in-progress out there. On the other hand, they might have simply needed to produce more options by themselves because no partner providing other possibilities was present. Garrod and Pickering (2013) noted that "taking part in a conversation is more straightforward than speaking or listening in isolation." In psycholinguistic models of speech production easier means *more automatized* and *less controlled*. Within the dual route account, the more straightforward pathway is retrieval. Our results with more non-canonical productions of the target items could be interpreted as follows: Monologues correspond (at least in our experimental setup) with more controlled speech productions — they might employ the assembly route — the participants think more about what they say — this gives them more options in speech production to become *creative*.

We observe a higher proportion of low-frequency target syllables produced as novel forms, while a higher proportion of high-frequency target syllables are produced in accordance with canonical patterns. This syllable-frequency effect clearly supports our expectation that low-frequency syllables are more susceptible to creative processes than high-frequency syllables. If there are certain patterns to these novel phonetic forms, e.g., if there is a difference in what kind of novel forms are produced from high- and low-frequency syllables, will be focused on in further research. Our analyses contribute to the understanding of speech production in different communicative settings and serve as a testbed for psycholinguistic models.

Acknowledgments

This research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — CRC-1646, project number 512393437, project A02.

References

Sana Tariq Ahmed and Gregory J. Feist. 2021. The Language of Creativity: Validating Linguistic Analysis to Assess Creative Scientists and Artists. *Frontiers in Psychology*, 12:724083.

Rachel Baker and Valerie Hazan. 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, 43(3):761–770.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models

Using lme4. *Journal of Statistical Software*, 67(1):1–48.

Simon Betz, Nataliya Bryhadyr, Olcay Türk, and Petra Wagner. 2023. Cognitive Load Increases Spoken and Gestural Hesitation Frequency. *Languages*, 8(1):71.

Paul Boersma and David Weenink. 2025. Praat: doing phonetics by computer.

Ronald Carter. 2015. *Language and Creativity: The art of common talk*, 1 edition. Routledge. First edition published by Routledge 2004.

Joana Cholin, Gary S. Dell, and Willem J. M. Levelt. 2011. Planning and articulation in incremental word production: Syllable-frequency effects in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1):109–122.

Daniel Duran, Leonie Schade, Joana Cholin, and Petra Wagner. 2025. Testing the strategic elicitation of creative pronunciations in monologues and dialogues. In *Elektronische Sprachsignalverarbeitung 2025 – Tagungsband der 36. Konferenz; Halle/Saale*, pages 85–92, Dresden. TUDpress.

Mirjam Ernestus and Natasha Warner. 2011. An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39(3):253–260. Editorial.

Simon Garrod and Martin J. Pickering. 2007. Automaticity of language production in monologue and dialogue. In Antje Meyer, Linda Wheeldon, and Andrea Krott, editors, *Automaticity and Control in Language Processing*, pages 1–20. Psychology Press.

Simon Garrod and Martin J. Pickering. 2013. Dialogue: Interactive Alignment and Its Implications for Language Learning and Language Change. In P.-M. Binder and K. Smith, editors, *The Language Phenomenon: Human Communication from Milliseconds to Millennia*, pages 47–64. Springer.

Mattias Heldner, Jens Edlund, Anna Hjalmarsson, and Kornel Laskowski. 2011. Very short utterances and timing in turn-taking. In *Interspeech 2011*, pages 2837–2840. ISCA.

Matthias Jilka. 2009. Assessment of phonetic ability. In Grzegorz Dogil and Susanne Maria Reiterer, editors, *Language Talent and Brain Activity*, pages 17–66. Mouton de Gruyter, Berlin / New York.

Michal Jirásek and František Sudzina. 2020. Big Five Personality Traits and Creativity. *Quality Innovation Prosperity*, 24(3):90–105.

Thomas Kisler, Uwe Reichel, and Florian Schiel. 2017. Multilingual processing of speech via web services. *Computer Speech & Language*, 45:326–347.

Anna K. Kuhlen and Rasha Abdel Rahman. 2017. Having a task partner affects lexical retrieval: Spoken word production in shared task settings. *Cognition*, 166:94–106.

Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13).

Russell V. Lenth. 2022. emmeans: Estimated Marginal Means, aka Least-Squares Means.

Willem J. Levelt, Ardi Roelofs, and Antje S. Meyer. 1999. A theory of lexical access in speech production. *The Behavioral and brain sciences*, 22(1):1–75.

Rense Nieuwenhuis, Manfred Te Grotenhuis, and Ben Pelzer. 2012. influence.ME: Tools for Detecting Influential Data in Mixed Effects Models. *R Journal*, 4(2):38–47.

Paul Paulus. 2000. Groups, teams, and creativity: The creative potential of idea-generating groups. *Applied psychology*, 49(2):237–262.

Beatrice Rammstedt, Christoph J. Kemper, Mira Céline Klein, Constanze Beierlein, and Anastassyia Kovaleva. 2014. Big Five Inventory (BFI-10). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. ZIS-GESIS Leibniz Institute for the Social Sciences. Version: 1.0.

Barbara Samlowski. 2016. *The syllable as a processing unit in speech production: Evidence from frequency effects on coarticulation*. Doctoral Dissertation, Universität Bielefeld, Bielefeld.

Petra Wagner, Marcin Włodarczak, Hendrik Buschmeier, Olcay Türk, and Emer Gilmartin. 2024. Turn-taking dynamics across different phases of explanatory dialogues. In *Proceedings of the 28th Workshop on the Semantics and Pragmatics of Dialogue*, pages 6–14.

Petra Wagner, Sina Zarrieß, and Joana Cholin. 2021. Effects of Time Pressure and Spontaneity on Phonotactic Innovations in German Dialogues. In *Interspeech 2021*, pages 3335–3339.

Michael Walsh, Bernd Möbius, Travis Wade, and Hinrich Schütze. 2010. Multilevel Exemplar Theory. *Cognitive Science*, 34(4):537–582.

Britta Wrede and Elizabeth Shriberg. 2003. Spotting "hot spots" in meetings: human judgments and prosodic cues. In *Proceedings of INTERSPEECH*, pages 2805–2808.

A Appendix

A.1 LMER models (Section 3)

term	Estimate	SE	df	t	Pr	
(Intercept)	1.723	0.54	23	3.21	0.004	**
task[DPX]	0.606	0.4	3904	1.52	0.129	
task[PND]	0.461	0.43	5363	1.06	0.288	
task[PWO]	-0.425	0.39	5810	-1.08	0.282	
task[DBI]	-1.431	0.47	7233	-3.02	0.003	**
fam.[peers]	0.25	0.15	23	1.62	0.118	
PNM[second]	0.202	0.18	34	1.12	0.273	
extraversion	0.149	0.08	21	1.93	0.067	
agreeableness	-0.088	0.09	23	-0.97	0.341	
neuroticism	-0.215	0.1	41	-2.05	0.047	*
openness	-0.067	0.08	17	-0.83	0.416	
task[DPX]:fam.[peers]	-1.018	0.12	3775	-8.67	<0.001	****
task[PND]:fam.[peers]	-0.617	0.13	6010	-4.71	<0.001	****
task[PWO]:fam.[peers]	-0.301	0.11	5837	-2.71	0.007	**
task[DBI]:fam.[peers]	-0.098	0.13	6883	-0.74	0.458	
task[DPX]:PNM[second]	-1.11	0.15	3369	-7.43	<0.001	****
task[PND]:PNM[second]	-0.815	0.16	4867	-5.09	<0.001	****
task[PWO]:PNM[second]	-0.332	0.14	3688	-2.36	0.018	*
task[DBI]:PNM[second]	-0.433	0.16	5916	-2.71	0.007	**
task[DPX]:extraversion	-0.371	0.05	3056	-6.79	<0.001	****
task[PND]:extraversion	-0.386	0.06	4660	-6.43	<0.001	****
task[PWO]:extraversion	-0.108	0.05	4984	-1.97	0.049	*
task[DBI]:extraversion	0.186	0.07	7289	2.63	0.009	**
task[DPX]:agreeableness	0.039	0.06	5766	0.65	0.517	
task[PND]:agreeableness	-0.006	0.07	6367	-0.08	0.933	
task[PWO]:agreeableness	0.207	0.07	5907	3.18	0.001	**
task[DBI]:agreeableness	0.038	0.08	7268	0.46	0.647	
task[DPX]:neuroticism	0.563	0.08	2824	6.64	<0.001	****
task[PND]:neuroticism	0.467	0.09	3839	5.16	<0.001	****
task[PWO]:neuroticism	0.128	0.09	2902	1.49	0.137	
task[DBI]:neuroticism	0.268	0.1	5920	2.79	0.005	**
task[DPX]:openness	-0.112	0.05	4967	-2.1	0.036	*
task[PND]:openness	-0.018	0.06	6361	-0.31	0.753	
task[PWO]:openness	0.004	0.05	6368	0.08	0.935	
task[DBI]:openness	-0.091	0.07	7401	-1.28	0.202	

Random effects. Number of obs: 7511, groups: speaker, 21

Groups	Name	Variance	Std.Dev.	Corr
speaker	(Intercept)	0.048	0.22	
Residual		0.881	0.939	

Table 4: LMER fixed effects coefficients and random effects of the **pause duration** model. Formula (following R notation according to the *lme4* package, see Bates et al., 2015): *pause.dur ~ task + familiarity + PNM + extraversion + agreeableness + neuroticism + openness + task:familiarity + task:PNM + task:extraversion + task:agreeableness + task:neuroticism + task:openness + (1|speaker).*

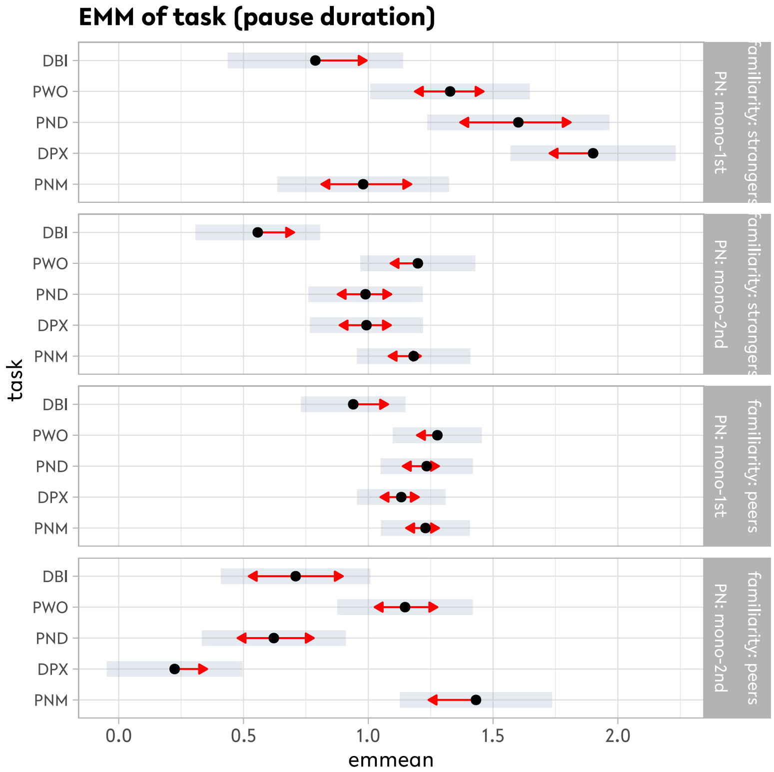

Figure 11: Estimated marginal means of *task* by *familiarity* and product naming order (PN) with the **pause duration** model. Shaded areas indicate confidence intervals, arrows show comparisons, reflecting “as much as possible the significance of the comparison of the two estimates” (Lenth, 2022).

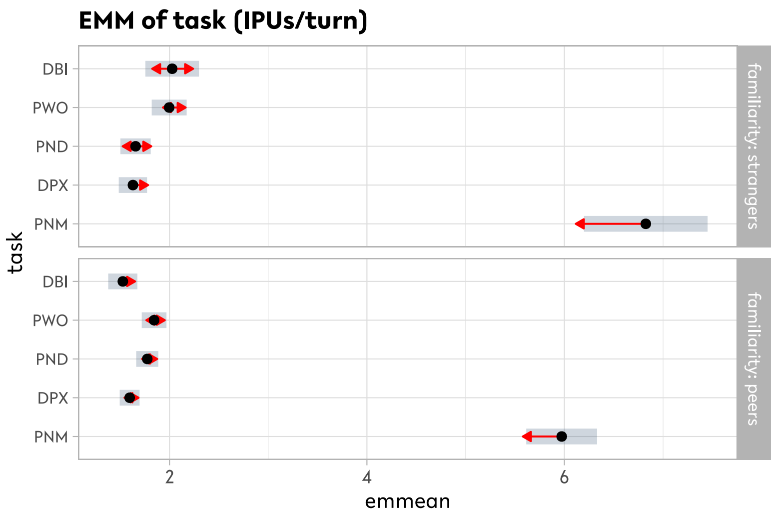

Figure 12: Estimated marginal means of *task* by *familiarity* with the **IPU/turn** model.

term	Estimate	SE	df	t	Pr	
(Intercept)	5.083	1.26	2888	4.03	<0.001	****
task[DPX]	-3.948	1.26	4643	-3.15	0.002	**
task[PND]	-3.223	1.26	4673	-2.55	0.011	*
task[PWO]	-3.425	1.27	5073	-2.69	0.007	**
task[DBI]	-2.899	1.3	4813	-2.23	0.026	*
fam.[peers]	-0.852	0.38	3468	-2.26	0.024	*
conscientiousness	1.04	0.18	1084	5.91	<0.001	****
neuroticism	-0.198	0.2	2149	-1.01	0.313	
openness	-0.355	0.29	5066	-1.23	0.217	
task[DPX]:fam.[peers]	0.819	0.37	5676	2.2	0.028	*
task[PND]:fam.[peers]	0.97	0.37	5679	2.59	0.01	**
task[PWO]:fam.[peers]	0.699	0.38	5679	1.85	0.065	
task[DBI]:fam.[peers]	0.352	0.39	5680	0.89	0.373	
task[DPX]:conscientiousness	-0.937	0.17	5664	-5.53	<0.001	****
task[PND]:conscientiousness	-1.194	0.17	5662	-6.87	<0.001	****
task[PWO]:conscientiousness	-0.965	0.18	5680	-5.39	<0.001	****
task[DBI]:conscientiousness	-1.006	0.18	5675	-5.72	<0.001	****
task[DPX]:neuroticism	0.282	0.2	3266	1.43	0.152	
task[PND]:neuroticism	0.146	0.2	3310	0.74	0.46	
task[PWO]:neuroticism	0.09	0.2	4078	0.45	0.651	
task[DBI]:neuroticism	0.107	0.21	3594	0.52	0.602	
task[DPX]:openness	0.324	0.29	5679	1.13	0.258	
task[PND]:openness	0.494	0.29	5680	1.72	0.086	
task[PWO]:openness	0.472	0.29	5679	1.64	0.102	
task[DBI]:openness	0.355	0.29	5679	1.22	0.221	

Random effects. Number of obs: 5680, groups: speaker, 21

Groups	Name	Variance	Std.Dev.	Corr
speaker	(Intercept)	0.023	0.15	
Residual		1.039	1.019	

Table 5: LMER fixed effects coefficients and random effects of **IPUs/turn** model. Formula: *n.IPU ~ task + familiarity + conscientiousness + neuroticism + openness + task:familiarity + task:conscientiousness + task:neuroticism + task:openness + (1|speaker)*

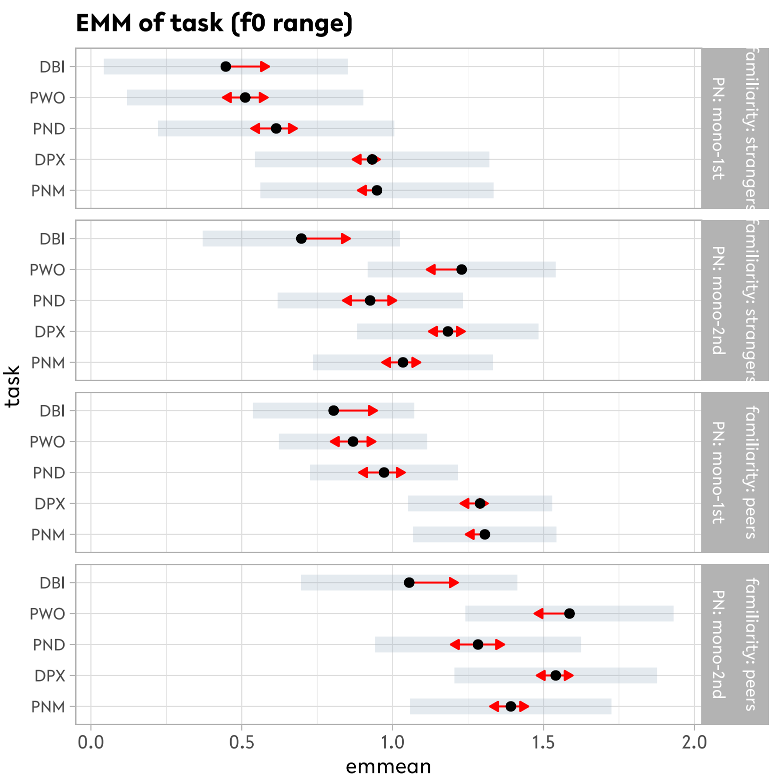

Figure 13: Estimated marginal means of *task* by product naming order (*PN*) and *familiarity* with the **f0 range** model.

term	Estimate	SE	df	t	Pr	
(Intercept)	-0.538	0.61	14	-0.88	0.391	
task[DPX]	1.361	0.19	14818	7.34	<0.001	****
task[PND]	0.563	0.24	14817	2.39	0.017	*
task[PWO]	0.33	0.24	14812	1.38	0.169	
task[DBI]	0.098	0.31	14812	0.31	0.755	
conscientiousness	0.194	0.12	15	1.58	0.136	
openness	-0.013	0.09	14	-0.14	0.892	
neuroticism	0.167	0.09	14	1.77	0.097	
extraversion	0.109	0.09	14	1.17	0.263	
PNM[second]	0.086	0.19	14	0.45	0.663	
fam.[peers]	0.357	0.17	14	2.06	0.059	
task[DPX]:conscientiousness	-0.414	0.04	14795	-10.22	<0.001	****
task[PND]:conscientiousness	-0.201	0.05	14816	-3.85	<0.001	***
task[PWO]:conscientiousness	-0.179	0.05	14817	-3.37	<0.001	***
task[DBI]:conscientiousness	-0.077	0.07	14814	-1.14	0.256	
task[DPX]:openness	0.282	0.03	14810	9.8	<0.001	****
task[PND]:openness	0.145	0.04	14817	4.13	<0.001	****
task[PWO]:openness	0.154	0.04	14818	4.13	<0.001	****
task[DBI]:openness	0.12	0.05	14816	2.33	0.020	*
task[DPX]:neuroticism	-0.187	0.03	14617	-5.72	<0.001	****
task[PND]:neuroticism	-0.263	0.04	14745	-6.35	<0.001	****
task[PWO]:neuroticism	-0.427	0.05	14816	-8.44	<0.001	****
task[DBI]:neuroticism	-0.354	0.07	14816	-5.36	<0.001	****
task[DPX]:extraversion	-0.113	0.03	14609	-3.76	<0.001	***
task[PND]:extraversion	0.027	0.04	14746	0.74	0.459	
task[PWO]:extraversion	0.188	0.04	14811	4.63	<0.001	****
task[DBI]:extraversion	0.097	0.05	14816	1.9	0.058	
task[DPX]:PNM[second]	0.165	0.07	14546	2.46	0.014	*
task[PND]:PNM[second]	0.225	0.09	14692	2.64	0.008	**
task[PWO]:PNM[second]	0.631	0.09	14797	6.65	<0.001	****
task[DBI]:PNM[second]	0.164	0.14	14817	1.2	0.230	

Random effects. Number of obs: 14849, groups: speaker, 21

Groups	Name	Variance	Std.Dev.	Corr
speaker	(Intercept)	0.092	0.303	
Residual		1.249	1.118	

Table 6: LMER fixed effects coefficients and random effects of **f0 range** model. Formula: *f0.range.z ~ task + familiarity + PNM + extraversion + conscientiousness + neuroticism + openness + task:conscientiousness + task:openness + task:neuroticism + task:extraversion + task:PNM + (1|speaker).*

term	Estimate	SE	df	t	Pr
(Intercept)	-1.335	0.46	18	-2.92	0.009 **
task[DPX]	1.829	0.29	3850	6.33	<0.001 ****
task[PND]	1.303	0.34	4687	3.82	<0.001 ***
task[PWO]	0.782	0.37	5077	2.13	0.033 *
task[DBI]	0.755	0.41	5366	1.85	0.064
conscientiousness	0.174	0.09	17	1.9	0.074
fam.[peers]	0.544	0.14	21	3.84	<0.001 ***
neuroticism	0.112	0.11	42	1.05	0.298
final[fin]	-0.299	0.11	5533	-2.7	0.007 **
openness	0.142	0.06	11	2.29	0.042 *
PNM[second]	-0.083	0.19	31	-0.43	0.671
extraversion	0.089	0.08	20	1.18	0.251
task[DPX]:conscientiousness	-0.388	0.06	4757	-6.84	<0.001 ****
task[PND]:conscientiousness	-0.193	0.07	5281	-2.62	0.009 **
task[PWO]:conscientiousness	-0.118	0.07	5284	-1.62	0.106
task[DBI]:conscientiousness	-0.039	0.09	5463	-0.43	0.665
task[DPX]:fam.[peers]	-0.256	0.09	4910	-2.7	0.007 **
task[PND]:fam.[peers]	-0.394	0.11	5475	-3.64	<0.001 ***
task[PWO]:fam.[peers]	-0.323	0.12	5077	-2.73	0.006 **
task[DBI]:fam.[peers]	-0.384	0.15	5519	-2.5	0.013 *
task[DPX]:neuroticism	-0.036	0.09	329	-0.39	0.693
task[PND]:neuroticism	-0.253	0.1	390	-2.64	0.009 **
task[PWO]:neuroticism	-0.302	0.1	492	-3.08	0.002 **
task[DBI]:neuroticism	-0.278	0.11	815	-2.61	0.009 **
task[DPX]:final[fin]	0.504	0.12	5533	4.19	<0.001 ****
task[PND]:final[fin]	0.395	0.13	5533	3.02	0.003 **
task[PWO]:final[fin]	0.275	0.13	5533	2.06	0.04 *
task[DBI]:final[fin]	0.607	0.16	5527	3.75	<0.001 ***
task[DPX]:PNM[second]	0.292	0.16	393	1.85	0.065
task[PND]:PNM[second]	0.424	0.17	501	2.49	0.013 *
task[PWO]:PNM[second]	0.623	0.18	594	3.49	<0.001 ***
task[DBI]:PNM[second]	0.295	0.2	903	1.49	0.136
task[DPX]:extraversion	-0.054	0.06	1169	-0.93	0.354
task[PND]:extraversion	0.014	0.06	1794	0.21	0.832
task[PWO]:extraversion	0.154	0.07	3082	2.31	0.021 *
task[DBI]:extraversion	-0.01	0.09	5110	-0.11	0.911

Random effects. Number of obs: 5569, groups: speaker, 20

Groups	Name	Variance	Std.Dev.	Corr
speaker	(Intercept)	0.04	0.199	
Residual		0.912	0.955	

Table 7: LMER fixed effects coefficients and random effects of **end-IPU f0 range** model. Formula: *f0.range.z ~ task + familiarity + PNM + final + extraversion + conscientiousness + neuroticism + openness + task:familiarity + task:PNM + task:final + task:extraversion + task:conscientiousness + task:neuroticism + (1 | speaker)*

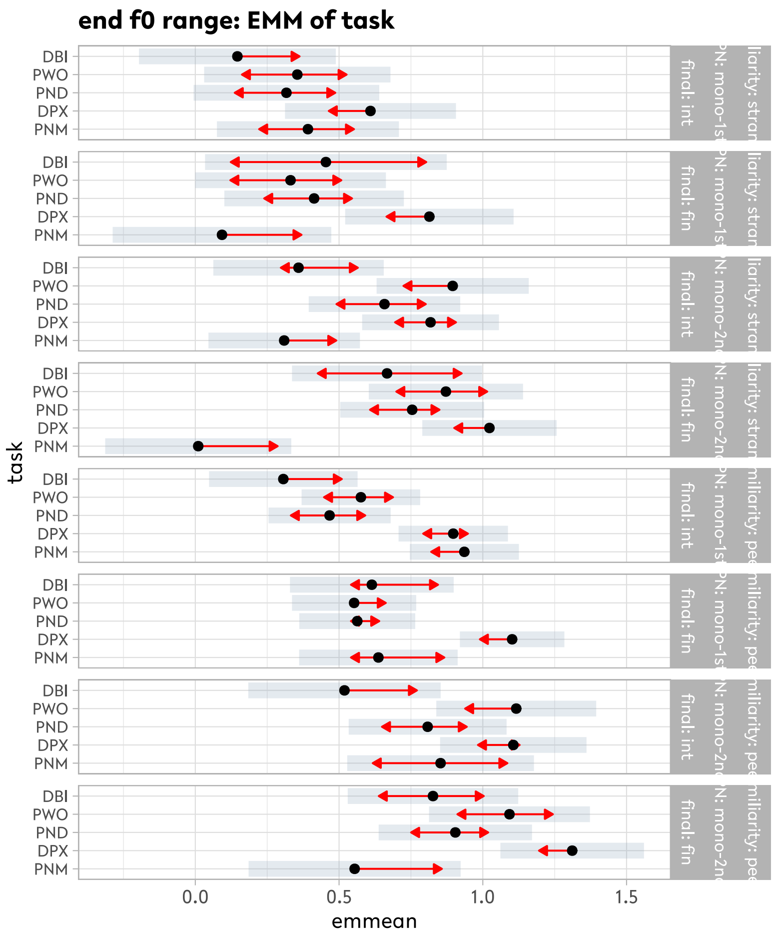

Figure 14: Estimated marginal means of *task* with the **end-IPU f0 range** model.

A.2 Additional Tables and Figures for Section 4

A.2.1 Novel Forms

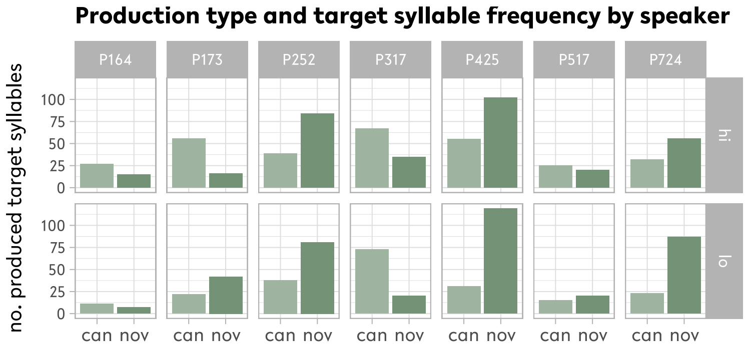

Figure 15: Number of uttered high- and low-frequency syllables within canonical and novel productions of each participant.

A.2.2 Monologues and Dialogues

setting & prod. type	P164	P173	P252	P317	P425	P517	P724	total	prod. type in settings (%)
monologue	35	68	188	112	162	37	127	729	(59.85)
can	19	32	52	58	35	9	25	230	31.55
nov	16	36	136	54	127	28	102	499	68.45
dialogue	25	68	54	83	145	43	71	489	(40.15)
can	19	46	25	82	51	31	30	284	58.08
nov	6	22	29	1	94	12	41	205	41.92
total	60	256	122	195	307	80	198	1218	
can	38	98	57	140	86	40	55	514	*42.20*
nov	22	158	65	55	221	40	143	704	*57.80*

Table 8: Distribution of novel and canonical productions of target syllables per participant within the monologue (PNM) and dialogue (PND & DPX) tasks.

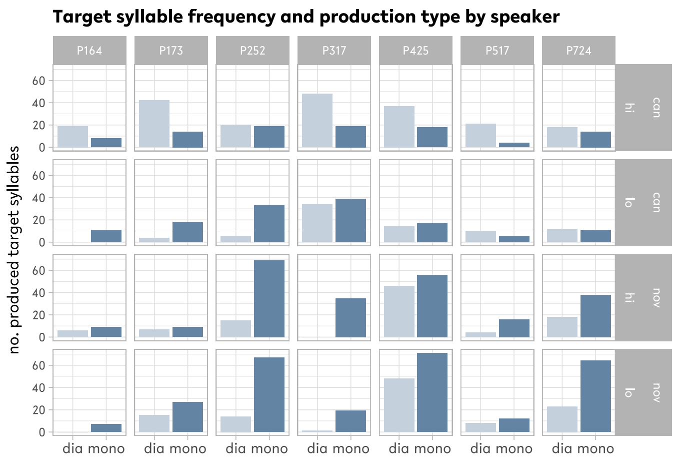

Figure 16: Number of uttered high- and low-frequency syllables within canonical and novel productions the of each participant in the monologue (PNM) and dialogue (PND & DPX) tasks.

A.2.3 Syllable Frequency

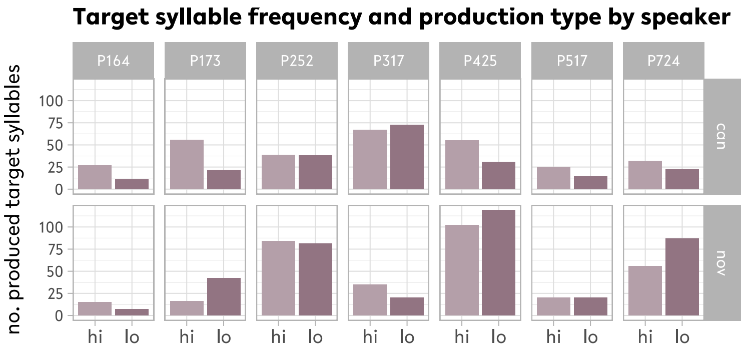

Figure 17: Number of uttered high- and low-frequency syllables within canonical and novel productions of each participant.

A.2.4 Product Naming Subset

setting/task & prod. type	P164	P173	P252	P317	P425	P517	P724	total	ratio prod. type in setting (%)
mono/PNM	35	68	188	112	162	37	127	729	(67.94)
can	19	32	52	58	35	9	25	230	31.55
nov	16	36	136	54	127	28	102	499	68.45
dia/PND	9	33	33	68	120	25	56	344	(32.06)
can	5	18	9	67	30	14	15	158	45.93
nov	4	15	24	1	90	11	41	186	54.07
total	44	101	221	180	282	62	183	1073	
can	24	50	61	125	65	25	40	388	*36.16*
nov	20	51	160	55	217	39	143	685	*63.84*

Table 9: Distribution of novel and canonical productions of target syllables per participant within the monologue and dialogue variations of the product naming task (PNM & PND).

prod. type & syl freq	P164	P173	P252	P317	P425	P517	P724	total	ratio syl frq in prod. type (%)
canonical	24	50	61	125	65	23	40	388	(36.16)
high	13	29	23	52	34	9	18	178	45.88
low	11	21	38	73	31	14	22	210	54.12
novel	20	51	160	55	217	39	143	685	(63.84)
high	13	14	82	35	98	19	56	317	46.28
low	7	37	78	20	119	20	87	368	53.72
total	44	101	221	180	282	62	183	1073	
high	26	43	105	87	132	28	74	495	*46.13*
low	18	58	116	93	150	34	109	578	*53.87*

Table 10: Distribution of high- and low-frequency target syllables per participant across novel and canonical productions within the monologue and dialogue variations of the product naming task (PNM & PND).

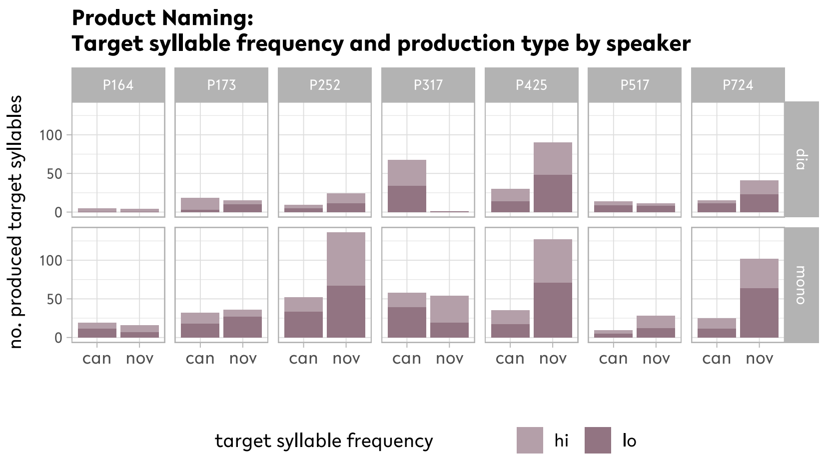

Figure 18: Number of uttered high- ("hi") and low- ("lo") frequency syllables within canonical and novel productions of each participant in the monologue and dialogue variations of the product naming task (PNM & PND).

A.3 GLMER models

term	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.3660	0.3837	-3.560	<0.001 ***
monodia[mono]	2.4125	0.3319	7.268	<0.001 ***
target_syl_freq[lo]	1.6079	0.3878	4.146	<0.001 ***
monodia[mono]:target_syl_freq[lo]	-2.1496	0.5003	-4.297	<0.001 ***

Random effects. Number of obs: 1218, groups: target_syl, 51; participant, 7

Groups	Name	Variance	Std. Dev.
target_syl	(Intercept)	0.5443	0.7378
participant	(Intercept)	0.5683	0.7539

Table 11: GLMER fixed effects coefficients and random effects of mono- & dialogue model ("model1.mdl"). Formula: *type_num ~ monodia * target_syl_freq + (1 | participant) + (1 | target_syl)*

term	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.7772	0.4581	-6.063	<0.001 ***
task[PND]	2.6496	0.4152	6.382	<0.001 ***
task[PNM]	3.5933	0.3944	9.110	<0.001 ***
target_syl_freq[lo]	4.0918	0.8792	4.654	<0.001 ***
task[PND]:target_syl_freq[lo]	-3.8822	0.9363	-4.146	<0.001 ***
task[PNM]:target_syl_freq[lo]	-4.3110	0.8890	-4.849	<0.001 ***

Random effects. Number of obs: 1218, groups: target_syl, 51; participant, 7

Groups	Name	Variance	Std. Dev.
target_syl	(Intercept)	0.2810	0.5301
participant	(Intercept)	0.5528	0.7435

Table 12: GLMER fixed effects coefficients and random effects of tasks model ("model2.mdl"). Formula: *type_num ~ task * target_syl_freq + (1 | participant) + (1 | target_syl)*

	npar	AIC	BIC	logLik	-2*log(L)	Chisq	Df	Pr(>Chisq)	
model1.mdl	6	1391.7	1422.3	-689.85	1379.7				
model2.mdl	8	1345.4	1386.2	-664.69	1329.4	50.301	2	1.194e-11	***

Table 13: Model comparison with ANOVA.

term	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.1537	0.4104	-0.374	0.70805	
monodia[mono]	1.0529	0.3612	2.915	0.00356	**
target_syl_freq[lo]	0.2108	0.3841	0.549	0.58320	
monodia[mono]:target_syl_freq[lo]	-0.5166	0.4865	-1.062	0.28826	

Random effects. Number of obs: 1073, groups: target_syl, 48; participant, 7

Groups	Name	Variance	Std. Dev.
target_syl	(Intercept)	0.3383	0.5816
participant	(Intercept)	0.6035	0.7769

Table 14: GLMER fixed effects coefficients and random effects of mono- & dialogue model on product naming subset. Formula: *type_num ~ monodia * target_syl_freq + (1 | participant) + (1 | target_syl)*