



Effects of time pressure and spontaneity on phonotactic innovations in German dialogues

Petra Wagner^{1,2}, Sina Zarriess^{1,2}, Joana Cholin¹

¹Faculty of Linguistics and Literary Studies, Bielefeld University, Germany

²Center for Cognitive Interaction Technology (CITEC), Bielefeld University, Germany

petra.wagner@uni-bielefeld.de, sina.zarriess@uni-bielefeld.de,
joana.cholin@uni-bielefeld.de

Abstract

Speech variation is often explained by speakers' balancing of production constraints (favoring phonetic reduction of high frequency, expected items) and listener orientation (favoring more canonical productions for low frequency, unexpected items). Less well understood are processes involving a structural re-organization of articulatory plans due to re-syllabification, e.g., resulting from processes involving massive reduction, epenthesis or metathesis. In this paper, we want to focus on two kinds of re-syllabifications: (1) within-system innovations, in which non-canonical forms occur, and (2) beyond-system inventions, which do not follow the phonotactic constraints of the language under consideration. We examine these processes in a corpus of spontaneous and read dyadic interactions of German, in which time pressure was controlled as an additional factor. Results show that spontaneity and time pressure will mostly lead to within-system innovations, favoring highly trained, unmarked articulatory routines, while minimizing information loss. However, occasionally speakers leave the beaten paths of highly trained articulatory routines, and invent novel phonotactic sequences which are at odds with the phonotactic grammar of German. Our results are discussed in the light of their implications for contemporary models of speech production.

Index Terms: re-syllabification, speech production, non-canonical, phonotactics, innovation

1. Introduction

Fluent speech production shows a high degree of deviation from what is often termed the "canonical form". The amount of this variation has been modeled by researchers as a function of entropy and surprisal (cf. [1] for a recent overview) as well as speaking style related factors, which have an impact on the speakers' choice of production variants [2, 3, inter alia]. Consequently, speaking style factors need to be taken seriously in theories and models explaining surface variation [4]. However, while speech production models have addressed the presence and absence of frequency related effects on syllable production [5, 6], their integration with speaking or interaction style related factors have hitherto not been taken into account, and many open questions with respect to the interplay between symbolic planning and articulatory realization remain [7, 3]. This is the more striking, as recent accounts of phonetic reduction have found effects of both listener- and talker-orientation to play a role [8].

Besides processes such as phonetic reduction, deletion or epenthesis on the level of the segment, "massive reduction", e.g., the deletion of one or several syllables in speech production, has been acknowledged to be present in spontaneous speech [9].

In this paper, we distinguish the following phenomena, which we illustrate using realistic examples of the corpus analyzed below:

- *reduction*: leading to "weaker" segmental realizations, e.g. *aber* ('but'):
/a:.bɛ/ → [a:.vɛ]
- *phone deletion*: fewer (audible) phones than in canonical form, e.g.: *eigentlich* ('actually'):
/ai.gɪt.liç/ → [ai.ŋ.liç]
- *phone/syllable epenthesis/insertion*: adding phones to canonical form, which may lead to additional syllables, e.g. *drei* ('three'):
/dʁai/ → [dɔ.ʁai]
- "massive" *syllable deletion*: fewer (audible) syllables than in canonical form involving re-syllabification, e.g. *eigentlich* ('actually'):
/ai.gɪt.liç/ → [aŋ.kliç]
- *re-syllabification*: new alignment of segments within the syllabic frame, e.g. *Flughafen* ('airport'):
/flu:k.ha:fn/ → [flu:k^h a:fn]

Many studies find that most reduction phenomena are non-categorical, and true deletions, where segments or entire syllables are fully removed out of the articulatory plan, may be a relatively rare phenomenon [10]. However, some studies have found evidence for categorical deletions, i.e., deletions of entire segments or syllables [11], and corpus studies on conversational speech have found "massive" deletions of entire syllables to be present in roughly 5% of the words, and segmental deletions to be present in 20% of all syllables [9]. Another type of surface variation which involves processes of re-syllabification but has received sparse attention is (vowel) epenthesis, a process which adds new syllables to the articulatory plan.

One of the few approaches that have addressed syllable-level categorical processes is the surprisal-entropy model by [12]. Their model predicts a preference for vowel epenthesis, leading to highly trained, unmarked syllable structures (CV syllables) as a result of listener-oriented re-syllabification, and reduction/deletion, but also more variation (including marked and unmarked syllable structures) for low surprisal words. Their model does not make any predictions for an interplay between surprisal and speaking style related factors.

Another model which predicts (possible) categorical syllable deletion, is described in [13]. Based on physiological constraints such as a limit of temporal incompressibility of segments, they predict a categorical deletion of syllables out of the articulatory plan if the costs for articulatory production outweigh perceptual clarity demands. Thus, there is a prediction

made on the grounds of speech rate or contextual factors such as time pressure. As their model is only concerned with syllable durations, they do not make any claims with respect to the necessary articulatory re-organization which may be the result of syllable deletion - e.g., a nucleus may be deleted, but the “stray” consonants may become codas or onsets of the syllables remaining in the articulatory plan.

In fact, speech-tempo induced re-syllabification from VC-syllables into less marked CV-syllables has been noticed long ago, and have been found to be frequently occurring in spontaneous speech [14]. However, its may be the result of perception rather than production [15]. Still, even if these re-syllabifications should only exist on the side of the listeners, they may be reproduced and initiate sound change [12, 16].

In our study, we want to further test the hypothesis that various types of *phonotactic innovations* may be caused by re-syllabification as a result of differences in speech production planning (reading vs. spontaneous production) and contextual factors such as time pressure.

We furthermore postulate that our results are best explained within a speech production model that allows for both an online-assembly of segments into innovative phonotactic structures and a direct route to high frequency, unmarked articulatory routines as part of a mental syllabary [6].

1.1. Hypotheses

In line with previous research, we expect

1. **a higher amount of reduction and phone deletions in spontaneous as compared to read speech**, leading to fewer audible segments, a higher distance to canonical forms, and shorter syllable durations,
2. that **reductions are stronger under time pressure**, leading to a greater number of syllable deletions, in line with [13], more phone deletions and a higher distance to canonical forms,
3. that **speakers re-syllabify canonical syllable structures more often in spontaneous than in read speech**. These re-syllabifications should mostly lead to higher frequency syllables and **unmarked forms**, e.g., CV syllables which speakers may easily retrieve from their mental syllabary,
4. that **syllable insertions (as the result of vowel epenthesis) mostly occur in spontaneous speech, but not under time pressure**. We expect that these epenthetic syllables typically result in highly trained, unmarked syllable structures (which would involve re-syllabification, in line with predictions by [12], or fully vocalic syllables,
5. that syllable deletions may occasionally lead to **inventions of novel syllabic structures which go beyond the regular phonotactic constraints of the language**. We particularly expect these cases to occur in spontaneous speech produced under time pressure.

2. Methods

2.1. Corpus recordings

We employed a corpus of task based dyadic interactions between 12 native speakers of German (6f, 6m) and 1, always identical, confederate. All speakers spoke a near standard variety of German, with very few dialectal markers. Each speaker

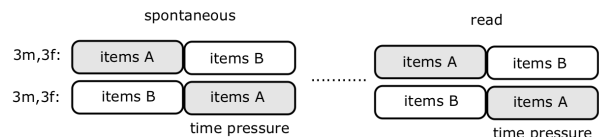


Figure 1: *Illustration of the recording setup. The x-axis represents time. Each speaker was either assigned to the group in which the items B were produced under time pressure, or the items A were produced under time pressure.*

was equipped with a set of materials typical for a tourist information belonging to a fictitious holiday resort (time tables, cultural programs, hiking trails, hotels etc.). The confederate acted as a tourist making a number of inquiries after, e.g., a hiking trail, movies for children, seasonal activities etc., thereby triggering a highly similar set of target words across all speakers in spontaneous interaction situations (e.g., a particular hotel name, time, street or bus stop). The target words were constructed in such a way that they covered the full German vowel inventory on several items. All recordings were made under laboratory conditions in a sound treated recording studio at University of Bonn, with separate close talking microphones, and with both interlocutors facing each other in the same room.

After half of the target words had been elicited by the confederate, she claimed to be in an extreme hurry, to create an overall atmosphere of time pressure. She continued to remind the speakers about her time pressure throughout the remainder of the interaction. The order of inquiry for target items was controlled as a between-subjects factor, so all target items were produced similarly often with and without time pressure. After the spontaneous interactions had been recorded, they were orthographically transcribed, and the same speakers were recorded in the identical interaction with the confederate, but this time both interlocutors read out aloud their respective turns based on the orthographic transcripts. Figure 1 illustrates the process and dimensions controlled during corpus recording.

2.2. Corpus annotations

All speakers’ productions were annotated manually on phone and syllable level using a narrow level of phonetic transcription. Subsequently, each syllable was also transcribed using the corresponding canonical form for the corresponding word. Canonical forms were determined based on the conventions described in [17] with a few systematic adaptations more in line with regular usage such as the uvular voiced fricative for /R/-realizations. We discarded utterance-final syllables due to their being particularly affected by phrase-final lengthening, which may influence syllable structure due to reasons that are beyond the current investigation. Cross-talk, breathing noises, pauses and disfluencies were also excluded the subsequent analyses. This leads to a total set of 20736 syllable tokens on which we based our analyses.

2.3. Syllable descriptions

2.3.1. Phonotactic structure

We parsed phonotactic structures of the narrowly annotated realizations into the following classes:

- open vs. closed
- with onset vs. onsetless
- vocalic vs. other

- consonantal vs. other

2.3.2. Deviation from canonical form

We calculated the Levenstein edit distance between canonical and realized syllable structures as an operationalization of deviations from canonical structures due to various phonological processes (reduction, deletion, epenthesis, metathesis).

2.3.3. Segmental length

We calculated the number of segments contained in realized syllables (as an indicator of segmental deletions in comparison with other styles).

2.3.4. Syllable duration

We obtained the duration of realized syllables (as an indicator of their level of reduction in comparison with other styles).

2.3.5. Syllable deletions and insertions

For each realized syllable, we subtracted the number of corresponding canonical syllables. This leads to negative numbers in case of deletions, and positive numbers in case of syllable insertions.

2.4. Analyses

2.4.1. Regression models

We calculated a series of linear mixed and generalized mixed regression models in order to determine the impact of the fixed factors *style* (spontaneous vs. read), *time pressure* as well as their interaction on the following dependent variables:

- syllable class (markedness)
- segmental length
- edit distance to canonical form
- syllable duration
- syllable epenthesis
- syllable insertion

Random factors (intercepts) across models were speaker and the canonical syllable. All analyses were performed with the statistical software R, version 4.0.2 [18], using the package *lme4* [19] for regression analyses. Significance of factors was determined with the *lmerTest* package [20] for the linear mixed models, and by a likelihood ratio model comparison for the generalized linear mixed models.

2.4.2. Analysing the well-formedness of phonotactic innovations

In order to test our hypothesis that speakers' massive reductions may lead to novel phonotactic structures beyond those belonging to the phonological grammar, we manually checked realized syllables in case of syllable deletions, and determined their phonotactic well-formedness.

3. Results

3.1. Reductions and deletions

As expected, spontaneous speech leads to syllable productions with a larger edit distance to canonical speech ($\beta = 4.652e - 02$, $se = 1.961e - 02$, $df = 2.555e + 04$, $t = 2.372$, $p =$

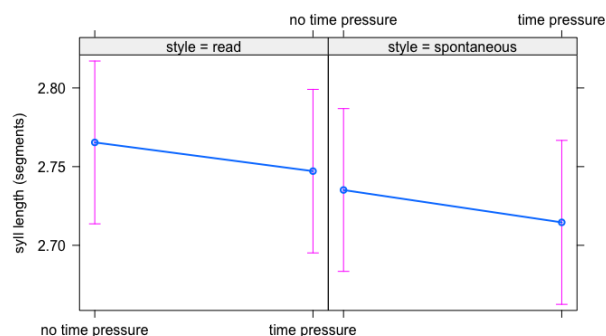


Figure 2: Illustration of syllable lengths (number of segments) across different styles (spontaneous, read) and under different time pressure conditions.

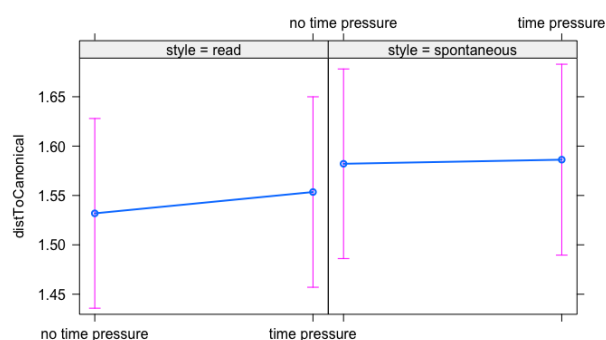


Figure 3: Illustration of Levenstein edit distances (based on segments) to canonical forms, across different styles (spontaneous, read) and under different time pressure conditions.

0.0177), and consists of fewer segments ($\beta = -3.860e - 02$, $se = 6.379e - 03$, $df = 2.413e + 04$, $t = -6.051$, $p = 1.46e - 09$), i.e., shows phone deletions (cf. Fig. 2). However, contrary to our expectations, spontaneous speech showed significantly longer syllable durations ($\beta = 6.574$, $se = 1.032$, $df = 24377.418$, $t = 6.372$, $p = 1.9e - 10$), possibly as a result of more prosodic phrase boundaries (which we did not control for).

Also contrary to our expectations, we did not find a main effect of time pressure on edit distance to canonical syllables (cf. Fig. 3). As expected, time pressure leads to fewer segments per syllable across speaking styles ($\beta = -1.825e - 02$, $se = 7.221e - 03$, $df = 2.395e + 04$, $t = -2.528$, $p = 0.0115$, cf. Fig. 2). This result goes hand in hand with the expected result that syllable durations are markedly shorter under time pressure ($\beta = -10.207$, $se = 1.108$, $df = 24528.840$, $t = -9.213$, $p < 2e - 16$).

3.2. Phonotactic structures

As expected, we see a higher probability for open syllables ($\beta = 0.05132$, $se = 0.02092$, $z = 2.453$, $p = 0.0142$) as well as syllables containing an onset ($\beta = 0.051908$, $se = 0.027018$, $z = 1.921$, $p = 0.0547$) in the spontaneous condition, indicating a production tendency for less marked syllable structures in spontaneous speech. Spontaneous speech

shows a lower probability for purely consonantal syllables ($\beta = -0.09682$, $se = 0.03559$, $z = -2.720$, $p = 0.00653$), but a higher probability for vocalic ones ($\beta = 0.1152$, $se = 0.03879$, $z = 2.972$, $p = 0.00296$).

Time pressure had no measurable effect on syllable structure preferences.

3.3. Syllable deletions and insertions

As expected, spontaneous speech led to a higher probability of massive reductions, e.g. syllable deletions ($\beta = 0.05813$, $se = 0.02930$, $z = 1.984$, $p = 0.0473$), but contrary to our expectation, this tendency was not notably increased by time pressure. No significant trends were found for syllable insertions.

3.4. Innovations beyond German phonotactics

Throughout the corpus, we detected occasional examples for phonotactic innovations which do not conform to the regular phonotactic constraints of German. These instances were very rare, and we refrain from statistical analyses due to their sparsity. The found cases could be classified as such:

- novel obstruent syllable nuclei: f, s, kç, tç, b, t
- novel obstruent-nasal onsets: tsm, km, xn
- novel obstruent-obstruent onsets: bz, db, jvk, kft, ft

Syllable codas showed no instances of innovative structures.

4. Discussion

This paper investigated under what circumstances speakers would use syllable productions that are indicative of 1) within-system innovations and 2) beyond-system inventions. Data were collected under conditions of spontaneous and read dyadic interactions with and without time pressure.

Spontaneous speech led to more deviations from a canonical form and a higher number of audible phones but also syllable deletions, which is in line with our H1. Unexpectedly, we also obtained longer syllable durations in this condition which might be due to a higher number of prosodic phrase boundaries and hence, phrase-final lengthening, as the result of online production planning. However, since we did not control for this, we refrain from any further interpretation of this result.

We also expected that time pressure would force speakers to resort to innovative phonetic sequences, during which phones and syllables are systematically reduced and deleted in order to accelerate production speed (H2). This, however, was only partially confirmed by our data: There was no main effect of time pressure on edit distance (indicating deviations from the canonical form) or syllable deletions, while we found fewer segments, i.e., more phone deletions under time pressure. Taken together, this finding is interesting as it is an insight into a production strategy which simplifies production effort by deleting material while sustaining a uniform level of edit distance to the canonical forms. This result provides support for Lindblom's H&H-theory [21] which predicts that speakers subtly balance production constraints and perceptual clarity. Also, our results provide further evidence that speech production and variation phenomena cannot be thoroughly explained without taking the cost of interaction into account.

H3 is fully supported by our results, as we found that speakers indeed prefer unmarked, open syllables with onsets when they are interacting in a spontaneous manner. We interpret these

findings as evidence for speakers' ability to perform innovative re-syllabifications of canonical forms, favoring syllable representation that are likely to be part of their mental syllabary, for which articulatory routines are readily available.

We could not find clear evidence for a higher probability of syllable insertions for spontaneous speech as predicted by our H4, which we would have expected as a result of listener orientation. However, we did find a preference for purely vocalic syllables in spontaneous speech, which may help a better perceptual distinction of highly informative consonantal segments, and which fits to the overall tendency for simple syllable structures in spontaneous speech.

As for H5, we refrain from any far-fetched interpretation in the light of very sparse data. Generally, however, we can confirm that speakers occasionally invent truly novel phonotactic sequences. Given the corpus size, however, we feel that our speakers' conformity to their phonotactic grammar in the vast majority of syllable productions is an interesting finding all by itself. Despite the high amount of reduction and overall variation that is so very characteristic for speech, speakers mostly adhere to the rules of their system. Whenever speakers deviate from canonical forms or constraints, they either invent new syllable nuclei (e.g., in the form of obstruents), or make up novel syllable onsets. We postulate that these instances of creative language use, albeit rare, may indeed initialize sound change [16]. Also, it would be interesting to see which articulatory routines and processes underlie these innovative forms.

We also find that our results are explicable within a model of speech production incorporating a dual route account for articulation planning, in which a mental syllabary provides access to ready-made articulatory routines for high-frequency syllables including reduced variants. While this retrieval route offers a short-cut to highly automatized, expected forms, the assembly route allows to build low-frequency and new forms from scratch. Against the background of the current results we would argue that the planning system opts for the retrieval route under conditions of casual spontaneous speech granting speakers a maximum of flexibility to be able to easily adapt to rapidly changing dialogue settings, prioritizing listener-oriented factors. The finding that non-canonical forms are easy to perceive in fluent connected speech supports this idea [11]. Under conditions of time pressure, speakers might uphold more canonical forms over highly reduced forms (as seen by maintaining edit distance), or construct novel syllables as a necessary result of massive reduction.

In spoken dialogue, speakers carefully juggle listener-oriented factors and production constraints to allow for smooth speech output as well as clear and effortless speech comprehension. Future research will target the circumstances under which speakers will resort to creative language use and what processes underlie the composition of these innovative forms.

5. Acknowledgements

We would like to thank Ralf Vogel for his inspiring thoughts on creative innovations and inventions within and beyond the grammatical system, and for ongoing discussions among the Bielefeld linguistics group investigating phenomena of creative language use. We would also like to thank Felicitas Haas and Julia Abresch for all the time they spent building the valuable corpus described above.

6. References

- [1] Z. Malisz, E. Brandt, B. Möbius, Y. M. Oh, and B. Andreeva, "Dimensions of segmental variability: Interaction of prosody and surprisal in six languages," *Frontiers in Communication*, vol. 3, p. 25, 2018.
- [2] B. Schuppler, M. Ernestus, O. Scharenborg, and L. Boves, "Acoustic reduction in conversational dutch: A quantitative analysis based on automatically generated segmental transcriptions," *Journal of Phonetics*, vol. 39, no. 1, pp. 96–109, 2011.
- [3] M. Ernestus, "Acoustic reduction and the roles of abstractions and exemplars in speech processing," *Lingua*, vol. 142, pp. 27–41, 2014.
- [4] P. Wagner, J. Trouvain, and F. Zimmerer, "In defense of stylistic diversity in speech research," *Journal of Phonetics*, vol. 48, pp. 1–12, 2015.
- [5] J. Cholin, W. J. Levelt, and N. O. Schiller, "Effects of syllable frequency in speech production," *Cognition*, vol. 99, no. 2, pp. 205–235, 2006.
- [6] J. Cholin, "The mental syllabary in speech production: An integration of different approaches and domains," *Aphasiology*, vol. 22, no. 11, pp. 1127–1141, 2008.
- [7] A. Bürki, M. C. Viebahn, I. Racine, C. Mabut, and E. Spinelli, "Intrinsic advantage for canonical forms in spoken word recognition: myth or reality?" *Language, Cognition and Neuroscience*, vol. 33, no. 4, pp. 494–511, 2018.
- [8] C. G. Clopper, R. Turnbull, F. Cangemi, M. Clayards, O. Niebuhr, B. Schuppler, and M. Zellers, "Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors," *Rethinking reduction*, pp. 25–72, 2018.
- [9] K. Johnson, "Massive reduction in conversational American English," in *Spontaneous speech: Data and analysis. Proceedings of the 1st session of the 10th international symposium*, 2004, pp. 29–54.
- [10] M. Ernestus and N. Warner, "An introduction to reduced pronunciation variants," *Journal of Phonetics*, vol. 39, no. S1, pp. 253–260, 2011.
- [11] A. Bürki, P. P. Cheneval, and M. Laganaro, "Do speakers have access to a mental syllabary? erp comparison of high frequency and novel syllable production," *Brain and Language*, vol. 150, pp. 90–102, 2015.
- [12] E. Hume and F. Mailhot, "The role of entropy and surprisal in phonologization and language change," *Origins of sound patterns: Approaches to phonologization*, pp. 29–47, 2013.
- [13] A. Windmann, J. Šimko, and P. Wagner, "Optimization-based modeling of speech timing," *Speech Communication*, vol. 74, pp. 76–92, 2015.
- [14] N. O. Schiller, A. S. Meyer, R. H. Baayen, and W. J. Levelt, "A comparison of lexeme and speech syllables in dutch," *Journal of Quantitative Linguistics*, vol. 3, no. 1, pp. 8–28, 1996.
- [15] K. J. De Jong, "Rate-induced resyllabification revisited," *Language and Speech*, vol. 44, no. 2, pp. 197–216, 2001.
- [16] J. J. Ohala, "Sound change as nature's speech perception experiment," *Speech Communication*, vol. 13, no. 1-2, pp. 155–161, 1993.
- [17] S. Kleiner and R. Knöbel, *Das Aussprachewörterbuch*, 7th ed., ser. Duden - Deutsche Sprache in 12 Bänden, Dudenredaktion, Ed., 2015, vol. 13.
- [18] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2020. [Online]. Available: <https://www.R-project.org/>
- [19] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
- [20] 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.
- [21] B. Lindblom, "Explaining phonetic variation: A sketch of the H&H theory," in *Speech production and speech modelling*. Springer, 1990, pp. 403–439.