



Effects of Lexical Class and Lemma Frequency on German Homographs

Barbara Samlowski¹, Petra Wagner¹, Bernd Möbius²

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

² Department of Computational Linguistics and Phonetics, Saarland University, Germany

{barbara.samlowski,petra.wagner}@uni-bielefeld.de, moebius@coli.uni-saarland.de

Abstract

German demonstrative pronouns, relative pronouns, and definite articles are segmentally identical but differ strongly in the frequency with which they appear. We examined the production of five such particles in a reading task. In a comparison of orthographically identical word pairs belonging to different lexical classes we found small but significant differences in word and vowel duration, prominence, and spectral similarity. Three of the particles in particular tended to be longer and more prominent when they occurred as demonstrative or relative articles than when they were assigned their usual role as definite articles.

Index Terms: lemma frequency, duration, prominence

1. Introduction

When speaking, people constantly have to navigate between their own wish to minimize muscular effort and the need to remain intelligible to their communication partner [1]. One possible strategy is to enunciate statistically improbable words and word combinations more carefully than common collocations. Studies of read speech as well of as spontaneous dialogues have shown that frequent words, independent of segmental length, tend to have shorter durations than rare ones [2-4]. However, a large part of the effect of word frequency can also be explained by other factors such as accentuation and prosodic boundaries [4] or conditional probability given the immediate context [3]. Content words already mentioned in the conversation also tend to be reduced in comparison with new occurrences [2, 3].

The exemplar theoretic approach integrates various influences on pronunciation into one model according to which numerous instances of words and word sequences are mentally stored in the form of concrete and highly detailed representations, the so-called exemplars [5]. Frequency information is encoded by having more exemplars stored for frequent words and word combinations than for rare ones. Recent exemplars are assumed to be more strongly activated than older ones. It is also possible that accented versus unaccented versions of a word and even different types of pitch accent are distinguished this way. As a recent study has shown, words occurring predominantly with one particular type of pitch accent are less varied in their pitch trajectories than words where the accent type is less predictable [6].

Frequency effects can also occur on the level of the syllable. An investigation of the influence of syllable frequency on pronunciation has demonstrated that duration values of frequent syllables are less predictable from their individual segments than duration values of rare syllables [7]. Frequent syllables also proved to have a stronger degree of coarticulation than rare syllables [8, 9]. A study by Pluymaekers and colleagues found word frequency effects on

syllable duration which were independent of syllable frequency by examining Dutch affixes in high and low frequency verbs [10].

Jurafsky and colleagues controlled syllable identity by examining English words which shared the same segmental form, but appeared in different grammatical roles [11]. This approach allowed them to explore whether frequency distinctions are made between different uses of the same word, i.e. whether there is not only an effect of word frequency but also of lemma frequency. Three of the four investigated word forms varied significantly in their degree of reduction depending on the grammatical role they played. Pure lemma frequency, however, did not prove to be a good predictor of these differences. A large part of the variance could equally well be explained by differences in predictability based on the surrounding word forms. Also, differences in lemma frequency did not necessarily correlate with how strongly words were reduced.

Many studies on the effect of word or lemma frequency are based on large corpora of spoken dialogues which were evaluated in a regression analysis [3, 4, 10, 11]. In the experiment described here, effects of word class and lemma frequency on pronunciation were examined in a controlled reading task, directly comparing segmentally identical homographs. In this way, a possible influence of syllable or word form frequency on the results is avoided.

2. Methods

2.1. Material

We investigated the German words "der" [de:ɐ], "die" [di:], "das" [das], "dem" [de:m], and "den" [de:n], which can be definite articles as well as relative or demonstrative pronouns. Which word is used in a concrete situation depends on the gender, number, and case of the noun to which it refers. While "den" and "das" are clearly attributed to one gender, "der", "die" and "dem" are more ambiguous: "der" can be masculine nominative as well as feminine genitive and dative, "die" can refer to plural as well as to feminine singular nouns, and "dem" is the dative particle for both masculine and neuter nouns.

Definite articles appear with high frequency in the German language, while relative or demonstrative pronouns are much less common. In the DeWaC corpus, a 1.5 billion word database of German web texts which was automatically tagged with information on lexical class and lemma form, about 90% of the ambiguous particles were categorized as definite articles, 7% as relative pronouns, and 3% as demonstrative pronouns. This frequency order generally applied to the individual particles as well, the only exception being the word "das", which had 71% occurrences as an

	[di:]	[de:ø]	[das]	[de:n]	[de:m]
DeWaC	1	4	15	29	68
Leipzig	1	3	13	22	42
Europarl	1	3	9	21	72
Subtlex	5	22	3	48	96
GF	2	9	1	48	91
Verbmobil	31	25	3	91	72

Table 1: Syllable ranks of the investigated particles in 6 databases

article, 22% as a demonstrative pronoun, and 7% as a relative pronoun [12].

As word forms, all five particles were highly frequent. In an extended corpus analysis, we found that they were among the 100 most common types in six databases of written and spoken language, both in terms of syllable frequency and in a comparison of phonological words [13]. In corpora which featured written or strongly planned language, the order of the particles in terms of frequency rank was "die", "der", "das", "den", and "dem", while databases containing spontaneous dialogues had "das" as the most frequent of the particles. Syllable frequency ranks for the six databases analyzed [12, 14-18] are shown in Table 1.

The present experiment was combined with a production study on word stress and acoustic prominence [19], so that the sentences from the two experiments could function as mutual distractors. Sentences with relative and demonstrative pronouns were formulated to match definite articles appearing in the stress study. For each set of three sentences composed accordingly, the half-syllables preceding and following the particle in question were kept constant. We constructed three such sets for both "den" and "das", and six each for "der" (masculine nominative and feminine dative), "die" (plural and feminine singular), and "dem" (masculine and neuter dative). All sentences were illustrated with pictures created with the help of the text-to-scene conversion program WordsEye [20]. In this way, test subjects were encouraged to think about the semantic content of the sentences.

2.2. Procedure

Illustrated sentences from both this experiment and the study on word and sentence stress [19] were placed in a quasi-random order and presented one at a time on a computer screen. Test subjects were asked to first look at each sentence and its illustration and then read the sentence aloud in a natural speaking style. They could proceed from one sentence to the next in a self-paced manner. Before the experiment, they were instructed to repeat a sentence if they noticed any slips of the tongue or other speaking errors. 30 native German speakers participated in the experiment, which was conducted in a sound treated chamber at Bielefeld University.

3. Results

To date, 20 test subjects (10 male, 10 female) have been analyzed. Sentences containing noticeable hesitations or restarts were omitted from the analysis, as were sentences

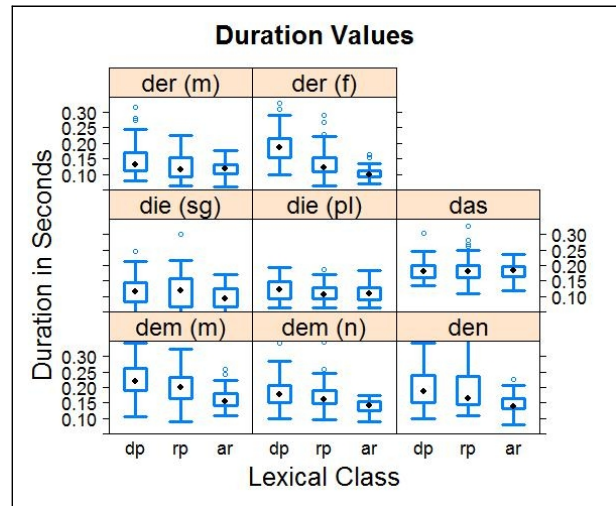


Figure 1: Duration values by words for demonstrative pronouns (dp), relative pronouns (rp), and articles (ar)

which were repeated by the test subjects due to speech errors. Slips of the tongue or reading errors which did not cause noticeable disfluencies were not removed unless they directly involved the investigated particle or its immediate context. In total, 1248 of 1440 stimuli were included in the analysis. Diagrams and statistical tests were made using the statistics program R [21]. Results for acoustic prominence and spectral similarity were not normally distributed, so that pairwise Wilcoxon rank tests were used to determine significance for these cases.

3.1. Duration

When comparing the duration values for the three grammatical categories across speakers, we found that demonstrative pronouns tended to be longer than relative pronouns, which in turn were longer than definite articles (see Figure 1). Differences between grammatical categories were particularly strong for feminine "der" and masculine "dem". For masculine "der", neutral "dem", and "den" the influence of lexical class was smaller, while hardly any differences were visible for the "das" and for singular or plural "die".

These results were generally confirmed in a direct comparison of words produced by the same speaker in the same context. Here, only feminine "der" and masculine "dem" showed clear differences between all three grammatical categories. Differences between categories were strongly reduced for masculine "der" and non-existent for "das" and singular/plural "die". For "den" and singular/plural "dem", relative pronouns tended to be closer in duration to demonstrative pronouns than to articles.

In a two-way analysis of variance of log-transformed word durations, significant effects were found for word identity ($p < .0001$, $F(7,1224)=107.8$) and lexical class ($p < .0001$, $F(2,1224)=81.5$), as well as for their interaction ($p < .0001$, $F(14,1224)=6.4$). Tukey HSD analyses confirmed differences between all three word classes ($p < .0001$). Word identity effects failed to reach significance for singular versus plural "die" ($p = .095$) and masculine versus feminine "der" ($p = .054$). Masculine and neuter "dem", however, were shown to be significantly different from each other ($p < .0001$). While all

particles except "das" and singular "die" contributed to the difference between demonstrative pronouns and articles ($p < .05$), the difference between relative pronouns and articles ($p < .05$), the difference between relative pronouns and articles was significant only for "den", feminine "der" and masculine "dem" ($p < .05$), and the effect of relative versus demonstrative pronouns was only significant for feminine "der" ($p < .0001$).

Examination of vowel durations revealed a pattern similar to the one in Figure 1. An analysis for log-transformed vowel duration confirmed main effects of word identity ($p < .0001$, $F(7,1224)=41.7$) and lexical class ($p < .0001$, $F(2,1224)=67.7$). The interaction between the two was also significant ($p < .0001$, $F(14,1224)=7.3$). According to Tukey HSD analyses, masculine versus neuter "dem" had significantly different vowel durations, as had masculine versus feminine "der" ($p < .05$). Contrary to the results for word durations, differences between demonstrative pronouns and articles failed to reach significance for singular "die" and masculine "der". The effect of demonstrative pronouns versus relative pronouns was only significant for feminine "der", the effect of relative pronouns versus articles only for masculine "dem" ($p < .0001$).

3.2. Acoustic Prominence

The degree of emphasis placed on demonstrative pronouns, relative pronouns, and articles may vary according to their specific function. We analyzed the particles using an automatic tagger which computes acoustic correlates of perceived prominence [22]. The tagger calculates a prominence value for each syllable nucleus of an utterance on the basis of pitch movement, overall intensity, nucleus duration, and spectral emphasis, i.e. the energy increase for higher spectrum frequencies. These factors are normalized across nuclei and weighted so as to maximize correlation with perceptual prominence ratings. In this study, only the immediately preceding and following syllables were given to the tagger as context. If the vowel preceding the target word in a particular sentence tended to be strongly reduced or deleted, the syllable before that was used for reference instead.

Across speakers and words, demonstrative pronouns proved to be significantly more prominent than relative pronouns or articles in pairwise Wilcoxon tests (Bonferroni-corrected, $p < .0001$). Relative pronouns had slightly lower prominence values than definite articles, although the difference was only significant in a direct comparison paired for word, speaker, and context ($V=25778$, $p < .01$). For "den", feminine "der", and masculine "dem", definite articles were found to have the lowest prominence values, while demonstrative pronouns were the most prominent of the three grammatical categories. The other words showed a different pattern, with relative pronouns being least prominent and only small differences appearing between demonstrative pronouns and articles.

3.3. Spectral Similarity

We also investigated spectral similarity between words belonging to different lexical classes, employing a method developed by Wade and colleagues [23] and Lewandowski [24]. This method compares word pairs by cross-correlating amplitude envelopes of the two words. Amplitude envelopes are computed for 4 frequency bands equally spaced on a logarithmic scale ranging from 80 to 7800 Herz, using a sampling rate of 500 Herz. The amplitude envelopes of the two words are then cross-correlated separately, and the

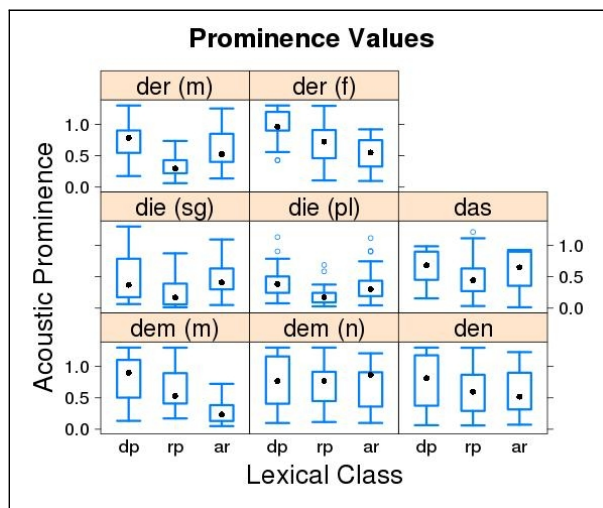


Figure 2: Prominence values by word for demonstrative pronouns (dp), relative pronouns (rp), and articles (ar)

highest resulting value is taken as a similarity measure. Amplitude profiles are representations that faithfully encode the speech signal as it unfolds over time without making specific assumptions about what types of cues might be extracted or which regions of the signal are the most important [23]. The similarity score also takes distortions in the temporal domain as one of several dimensions of acoustic distance into account.

Using this method, we determined the similarity of word pairs which were produced by the same speaker in the same phonological context, but belonged to different grammatical categories. In the analysis, demonstrative pronouns were found to be significantly closer to relative pronouns than to articles (pairwise Wilcoxon rank tests, Bonferroni-corrected, $p < .01$). This trend was followed by all particles apart from "das" and "die" (both singular and plural). We also examined the similarity of words within the three grammatical categories, comparing realizations by the same speaker in different phonological contexts, but found no systematic effects there.

4. Discussion

In the present experiment, segmentally identical demonstrative pronouns, relative pronouns, and articles were found to have significantly different vowel and word duration values. Demonstrative pronouns were longest, followed by relative pronouns. Articles had the shortest duration values. Although the differences correspond to differences in lemma frequency, with frequent lemmas being shorter, the effect may also be due to other factors. It seems plausible that definite articles were not only more common than pronouns, but were also better predictable from their context. The possibility of an effect of conditional probability is strengthened by the fact that the carrier sentences for relative and demonstrative pronouns were specifically constructed to match the articles in terms of phonetic context so that they may have been more highly contrived.

Apart from effects of lemma frequency or conditional probability, the comparatively long duration of demonstrative pronouns may also be due to their function as deictic expressions. In the study by Jurafsky and colleagues, even

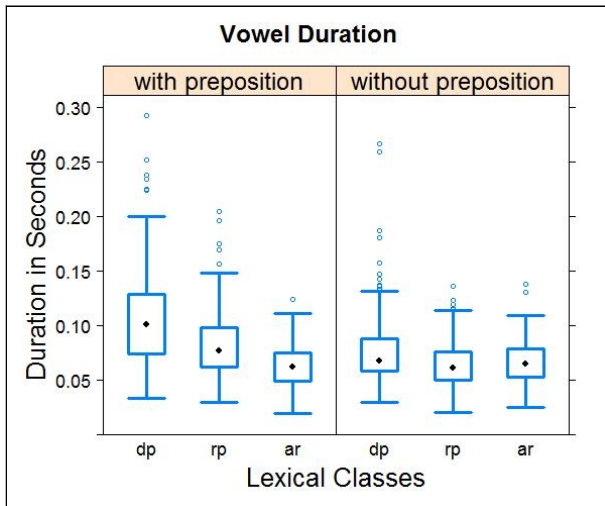


Figure 3: *Vowel duration values according to lexical class and preceding context (with or without preceding preposition)*

though the English particle "that" appeared most frequently as a demonstrative pronoun, it was shown to have longer duration values in this category compared to its use as a relative pronoun, complement, or determiner [11]. In our data, the highest acoustic prominence values were found for demonstrative pronouns, indicating that these were more strongly emphasized than other lexical classes.

Relative pronouns tended to be weakest in prominence, even though they were generally longer than articles. This result is probably not attributable to the particles themselves, but rather to the syntactic structure of the carrier sentences. As relative pronouns usually appear after phrase boundaries, the syllable preceding them may have a strongly increased prominence value due to final lengthening, which would lead to the relative pronoun being comparatively less prominent. In the carrier sentences used for "dem", feminine "der", as well as one of the sentences with "den", this problem was avoided by embedding the relative pronouns in prepositional phrases. For these particles the difference in prominence between relative pronouns and articles was reversed, or at least reduced (in the case of neuter "dem").

Interestingly, "dem", "den", and feminine "der" also showed stronger durational differences than the other particles. For these words, demonstrative pronouns and articles tended to be preceded by the same preposition as the matching relative pronouns, leading to a more highly controlled context and, consequently, to an increased level of ambiguity. Test subjects might have tried to counteract this ambiguity by increasing the duration differences between grammatical categories. Figure 3 compares grammatical categories with and without a preceding preposition. While articles remained comparatively unaffected, demonstrative and relative pronouns tended to be longer if they were preceded by a preposition. As vowel lengths were compared, the results were not influenced by a possible effect of phrase boundary on the closure duration of relative pronouns.

Word frequency might also have had an influence on how strongly individual particles were affected by the grammatical role in which they were used. Both "die" and "das", which showed the smallest differences between lexical classes, belong to the most frequent word and syllable types.

However, differences in word or syllable frequency do not explain why the particle "der", which is also very frequent, showed a rather strong effect of lexical class as a feminine dative, but not as a masculine nominative.

Frequency effects might play a role on a more differentiated level if "der" appeared more frequently as a masculine nominative than as a feminine dative particle, and if singular and plural "die" were similar in frequency. However, as we did not have any data on gender and case specific frequencies of the individual particles, we cannot verify this possibility. Most likely the differences between lexical classes and the varying extent to which they are reproduced in specific particles and contexts are due to a complex interaction between syntactic and prosodic factors as well as frequency and predictability effects on various hierarchical levels.

5. Conclusions

Of the particles analyzed, feminine "der", masculine "dem", and "den" showed the strongest durational differences between grammatical categories, with demonstrative pronouns being longer than relative pronouns, which in turn were longer than definite articles. While masculine "der" and neuter "den" had reduced differences in the same direction, practically no effect was found for "das" and singular or plural "die". Acoustic prominence increased with duration for feminine "der", masculine "dem" and "den". The particles "das", masculine "der", and singular and plural "die" had clearly reduced prominence levels for relative pronouns, even though there was no corresponding reduction of syllable or vowel duration. Here, the reduced prominence of relative pronouns was probably an artifact due to the syntactic structure of the sentences. As syllable prominence was calculated relative to the preceding and following syllable, relative pronouns were susceptible to final lengthening effects on the preceding syllable in cases where it was immediately followed a phrase boundary. In an investigation of spectral similarity, demonstrative pronouns were found to be more similar to relative pronouns than to articles.

The varying degree to which the individual particles reacted to changes in lexical class may partly be explained by frequency effects. The particles "die" and "das", which were the least affected by the grammatical category in which they were placed, belong to the most frequent word and syllable forms, particularly in spoken dialogues. There is also a strong possibility that the results were affected by the sentence structure and context in which the particles were placed, as particles that showed clear duration differences were predominantly embedded in prepositional phrases.

6. Acknowledgements

This study was funded by the German Research Foundation DFG (Priority Program 1234). Many thanks go to Natalie Lewandowski for providing us with the Matlab scripts used to analyze spectral similarity.

7. References

- [1] B. Lindblom, "Explaining phonetic variation: A sketch of the H&H theory," in *Speech production and speech modelling*, W. J. Hardcastle and A. Marchal, Eds., Dordrecht; Boston: Kluwer Academic Publishers, 1990, pp. 403-439.
- [2] R. E. Baker and A. R. Bradlow, "Variability in word duration as a function of probability, speech style, and prosody," *Language and Speech*, vol. 52, pp. 391-413, 2009.
- [3] A. Bell, J. M. Brenier, M. Gregory, C. Girand, and D. Jurafsky, "Predictability effects on durations of content and function words in conversational English," *Journal of Memory and Language*, vol. 60, pp. 92-111, 2009.
- [4] M. Aylett and A. Turk, "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," *Language and Speech*, vol. 47, pp. 31-56, March 2004.
- [5] J. B. Pierrehumbert, "Exemplar dynamics: Word frequency, lenition and contrast," in *Frequency effects and the emergence of linguistic structure*, J. L. Bybee and P. J. Hopper, Eds., Amsterdam: John Benjamins Publishing Company, 2001, pp. 137-157.
- [6] K. Schweitzer, S. Calhoun, H. Schütze, A. Schweitzer, and M. Walsh, "Relative frequency affects pitch accent realisation: Evidence for exemplar storage of prosody," in *Proceedings of the 13th Australasian International Conference on Speech Science and Technology*, Melbourne, 2010, pp. 62-65.
- [7] A. Schweitzer and B. Möbius, "Exemplar-based production of prosody: Evidence from segment and syllable durations," in *Proceedings of Speech Prosody 2004*, Nara, 2004, pp.459-462.
- [8] U. Benner, I. Flechsig, G. Dogil, and B. Möbius, "Coarticulatory resistance in a mental syllabary," in *Proceedings of the International Congress of Phonetic Sciences*, Saarbrücken, 2007, pp. 485-588.
- [9] K. Croot and K. Rastle, "Is there a syllabary containing stored articulatory plans for speech production in English?," in *Proceedings of the 10th Australian International Conference on Speech Science and Technology*, Sydney, 2004, pp. 376-381.
- [10] M. Pluymackers, M. Ernestus, and H. R. Baayen, "Lexical frequency and acoustic reduction in spoken Dutch," *Journal of the Acoustical Society of America*, vol. 118, pp. 2561-2569, October. 2005.
- [11] D. Jurafsky, A. Bell, and C. Girand, "The role of the lemma in form variation," in *Laboratory Phonology 7*, C. Gussenhoven and N. Warner, Eds., Berlin; New York: Mouton de Gruyter, 2000, pp. 3-34.
- [12] M. Baroni and A. Kilgarriff, "Large linguistically-processed web corpora for multiple languages," in *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, 2006, pp. 87-90.
- [13] B. Samlowski, B. Möbius, and P. Wagner, "Comparing syllable frequencies in corpora of written and spoken language," in *Proceedings of Interspeech 2011*, Florence, 2011, pp. 637-640.
- [14] U. Quasthoff, M. Richter, and C. Biemann, "Corpus portal for search in monolingual corpora," in *Proceedings of the 5th International Conference on Language Resources and Evaluation*, Genua, 2006, pp. 1799-1802.
- [15] P. Koehn, "Europarl: A parallel corpus for statistical machine translation," in *Proceedings of the 10th Machine Translation Summit*, Phuket, 2005, pp. 79-86.
- [16] M. Brysbaert, M. Buchmeier, M. Conrad, and A. M. Jacobs, "The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German," *Experimental Psychology*, vol. 58, pp. 412-424, July 2011.
- [17] S. Dickgießer. (2012). *Korpus GF, Gespräche Im Fernsehen: Talkshows, Diskussionen, Interviews*. Available: <http://agd.ids-mannheim.de/download/gfdok.pdf>
- [18] W. Wahlster, "Mobile speech-to-speech translation of spontaneous dialogs: An overview of the final Verbmobil system," in *Verbmobil: Foundations of speech-to-speech translation*, W. Wahlster, Ed., Berlin: Springer, 2000, pp. 3-21.
- [19] B. Samlowski, P. Wagner, and B. Möbius, "Disentangling lexical, morphological, syntactic and semantic influences on German prosody - Evidence from a production study," in *Proceedings of Interspeech 2012*, Portland, 2012, pp. 2406 – 2409.
- [20] B. Coyne and R. Sproat, "WordsEye: An Automatic Text-to-Scene Conversion System," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, Los Angeles, 2001, pp. 487-496.
- [21] R Development Core Team. (2010). *R: A language and environment for statistical computing*. Available: <http://www.r-project.org>
- [22] F. Tamburini and P. Wagner, "On automatic prominence detection for German," in *Proceedings of Interspeech 2007*, Antwerp, 2007, pp. 1809-1812.
- [23] T. Wade, G. Dogil, H. Schütze, M. Walsh, and B. Möbius, "Syllable frequency effects in a context-sensitive segment production model," *Journal of Phonetics*, vol. 38, pp. 227-239, 2010.
- [24] N. Lewandowski, "Talent in nonnative phonetic convergence," Ph.D. dissertation, Institut für Maschinelle Sprachverarbeitung, Stuttgart University, Stuttgart, Germany, 2011.