9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

# Impact of prosodic structure and information density on dynamic formant trajectories in German

*Erika Brandt, Frank Zimmerer, Bistra Andreeva, Bernd Möbius*

Saarland University, Saarbrücken, Germany
ebrandt@coli.uni-saarland.de

## Abstract

This study investigated the influence of prosodic structure and information density (ID), defined as contextual predictability, on vowel-inherent spectral change (VISC). We extracted formant measurements from the onset and offset of the vowels of a large German corpus of newspaper read speech. Vector length (VL), the Euclidean distance between F1 and F2 trajectory, and F1 and F2 slope, formant deltas of onset and offset relative to vowel duration, were calculated as measures of formant change. ID factors were word frequency and phoneme-based surprisal measures, while the prosodic factors contained global and local articulation rate, primary lexical stress, and prosodic boundary. We expected that vowels increased in spectral change when they were difficult to predict from the context, or stood in low-frequency words while controlling for known effects of prosodic structure. The ID effects were assumed to be modulated by prosodic factors to a certain extent. We confirmed our hypotheses for VL, and found expected independent effects of prosody and ID on F1 slope and F2 slope.

**Index Terms**: information density, prosodic modeling, vowel-inherent spectral change

## 1. Introduction

In recent years, there has been a number of studies investigating the relationship between information density (ID) and prosody, and both of their contributions to explaining phonetic variability. Here, information density is defined as the predictability of a linguistic unit in context. Studies have shown that vowels were longer and showed more spectral distinctiveness when they were difficult to predict from the context [1, 2].

ID has been interpreted as being mediated by prosodic structure in its effect on phonetic structure [3]. This strong version of the Smooth Signal Redundancy Hypothesis (SSRH) has been refuted by studies showing that phonetic reduction cannot be attributed entirely to prosodic features [4]. Others proposed that prosodic structure, such as pitch contour, were directly influenced by the predictability of the underlying linguistic structure [5]. In addition, studies have investigated the interaction of prosody and ID on local phonetic structures [6].

Vowel-inherent spectral change (VISC) measures spectral change based on the initial and final portion of the vowel [7, 8]. Taking two samples within the vowel functions as a way of time-normalization and enables comparisons between vowels of different duration. Among the proposed measures are F1 and F2 slope. For formant slope, the difference between initial and final formant frequency is set into relation to the duration of the vowel. It is calculated as

$$F_n\_Slope = \frac{\Delta F_n}{VowelDur}. \tag{1}$$

There are several other measures of dynamic formant trajectories expressing the relationship between equidistant, time-normalized formant measurements of vowels. In the F1/F2 plane, vector length (VL) can be interpreted as an indicator of the amount of formant change. This measure is expressed as the Euclidean distance between the onset and offset of F1 and F2 values. The longer the distance between these values the greater is the magnitude of change within the vowel. VL is calculated as

$$VL = \sqrt{((F1_i - F1_f)^2 + (F2_i - F2_f)^2)}. \tag{2}$$

Vowel dynamics can also be expressed by curve-fitting parameterizations, such as orthogonal polynomial, or discrete cosine transformation. Both methods describe the spectral curve by using a small set of coefficients for spectral mean, slope, and curvature of the formant trajectory. Simple VISC measures were performing equally well in distinguishing vowel categories as more complex curve-fitting spectral parameters [9].

VISC and local variation in pitch interact with one another. Regional varieties with greater VISC showed a more "exaggerated" F0 contour with earlier F0 rise and a steeper F0 fall within the vocalic nuclei carrying the pich accent in comparison to varieties with low VISC [10]. Systematic differences in VISC and F0 dynamics have been observed to help differentiate vowel identities in languages with large vowel inventories, e. g. Saterland Frisian [11] or Welsh [12].

At increased speech rate vowel targets are more centralized due to articulatory target undershoot. Identification ratings of vowels at fast speech rate were poorer than at normal speech rate [13]. In a more recent study, [14] found that listeners did not have difficulties identifying vowels at increased speech rate. Patterns of formant change were similar across fast and normal speech rate. Consonantal context was identified as the main predictor of different VISC patterns.

In languages with tense and lax vowels, such as German, acoustic analyses have identified distinct patterns for formant movement. Lax vowel production is characterized by a short target and slower release into the following consonant. Tense vowels usually involve longer duration with increased hold in target position, as well as rapid transitions moving into following consonants [15]. For German, [16] found distinct dynamic F1 formant trajectories that listeners can use to disambiguate tense and lax vowels.

This study investigated the impact of prosodic structure and information density on VISC measurements VL, F1 and F2 slope. We hypothesized that vowels displayed increased spectral change when they were difficult to predict from the context, or in low-frequency words. We also expected higher spectral change in vowels in primary lexical stress position than in unstressed position. VISC was expected to correlate negatively with speech rate acceleration, and positively with vowel duration and the occurrence of a prosodic boundary.

10.21437/SpeechProsody.2018-24

# 2. Methodology

## 2.1. Material

### 2.1.1. Speech corpus

We used the Siemens Synthesis Corpus (SI1000P) that contains 1000 newspaper sentences from the SI1000 newspaper corpus read by two professional male German native speakers [17]. The studio quality recordings were filtered and down-sampled from 48 kHz to 16 kHz. Forced-aligned segmentations using WebMAUS for German [18] were manually verified by phonetic experts.

### 2.1.2. Language modeling corpus

As a basis for the language models used in this study we processed the corpus SDeWaC which is a subset of the DeWaC corpus [19]. It contains 846,159,403 word tokens and 1,094,902 word types from .de web domains. Duplicates and passages which were deemed unsuitable for parsing were cleaned from the original DeWaC to create the subset. The grapheme-to-phoneme component of German Festival [20] was used to transcribe the corpus.

## 2.2. Data analysis

### 2.2.1. Formant measurements

Formant measurements were taken using the Praat [21] command ”To Formant (burg)” with the default values of time step 0, a maximum of five formants, the maximum formant value set at 5000 Hz for male speakers, an analysis window of 25 ms, and pre-emphasis from 50 Hz. Formant values from sampling points at initial and final position (20 % and 80 % of the vowel duration) were used to calculate the VISC measures.

A cleaning procedure was performed on the entire data set including vowels in function and content words. Data cleaning involved plotting the F1 and F2 values of the vowel phonemes per speaker with their respective ellipse at 95 % confidence interval to identify spurious values, as described in [22]. These data points were manually checked and excluded from the data depending on whether they were tracking errors. From originally 86,706 vowels only 0.73 % were excluded in the process. Formant values were normalized per speaker using Lobanov normalization which has outperformed other normalization procedures in comparative studies [23].

### 2.2.2. Language modeling

80 % of the SDeWaC corpus was used to create a training corpus for language modeling. The data was fed into SRILM [24] to calculate phoneme language models including word and sentence boundaries. The predictability output for biphone and triphone of the following and preceding context was then transferred into surprisal (Equation 3). We chose small n-phone sizes because we found they showed the strongest relationship to phonetic variability. We also extracted word frequency and phoneme probability from the SDeWaC training corpus. All ID measures were log-transformed because of positive skewness.

$$S(unit_i) = -log_2 P(unit_i|context). \tag{3}$$

### 2.2.3. Prosodic factors

We estimated global articulation rate on the sentence level and local articulation rate on the word level (phones per second). Both rate measures were mean-centered, separately for each speaker. In addition, we included primary lexical stress (levels: stressed, unstressed), and prosodic boundary information (levels: none, word, and phrase) as prosodic control factors.

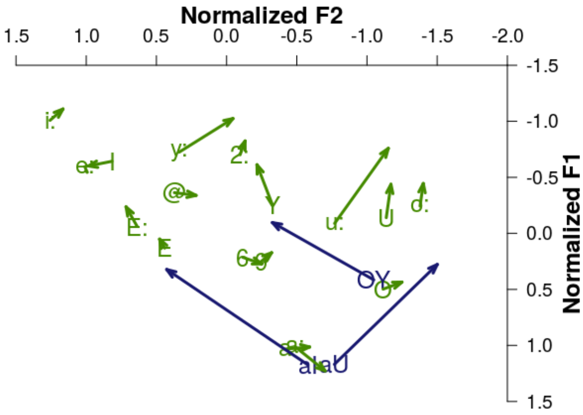

Figure 1: *Spectral change indicated by arrows from onset to offset in monophthongs and diphthongs (SAMPA). Diphthongs are displayed in blue, monophthongs in green.*

# 3. Results

Because word class and word frequency are known to correlate strongly we decided to exclude vowels in function words from the analysis [25]. Diphthongs inherently displayed more VISC than monophthongs (Figure 1). In order to have a more uniform data set we ran linear mixed-effects models (LMMs) only on monophthongs. In total, there were 57,728 monophthongs in content words in the corpus.

## 3.1. Linear mixed-effects models

Statistical analysis was performed using LMMs [26]. LMM structure was held constant across VISC measures to ensure comparisons between the models. Since we only found very low positive correlations for biphone surprisal of the preceding context (logBiSur) and VL, absolute F1 slope (absF1Slope) and F2 slope (absF2Slope), as well as for triphone surprisal of the preceding context (logTriSur) and VL, only LMMs for the preceding context were tested (Figure 2 for Pearson’s $r$). Surprisal values based on following context for biphone (logBiFolSur) or triphone (logBiFolSur) did not correlate positively with formant change measures. If surprisal had a significant effect on VISC, interaction models were calculated investigating potential interaction effects between surprisal and prosodic factors. The three prosodic factors used here, articulation rate, primary lexical stress and boundary, were entered separately as interaction terms in the models.

In order to avoid collinearity in the LMMs we performed a correlation analysis between the fixed effects prior to model training. There were low to moderate negative correlations between phoneme probability (PhProb) and surprisal values, which excluded phoneme probability from further modeling. In addition, there was a low positive correlation between global and local articulation rate ($r = 0.19$). Average vowel duration and local ($r = -0.34$) and global articulation rate ($r = -0.13$) correlated negatively. The factor primary lexical stress showed low positive correlations with word frequency ($r = 0.18$) and

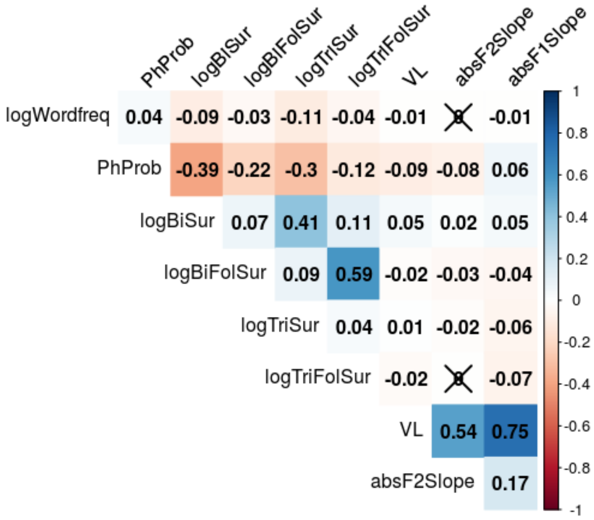

Figure 2: *Correlation matrix (Pearson's $r$) for VISC measures (VL, absolute formant slopes) and ID measures (surprisal, word frequency, and phoneme probability). Insignificant correlations at significance level 0.05 are crossed out. Positive correlations are displayed in blue shades and negative ones in red shades.*

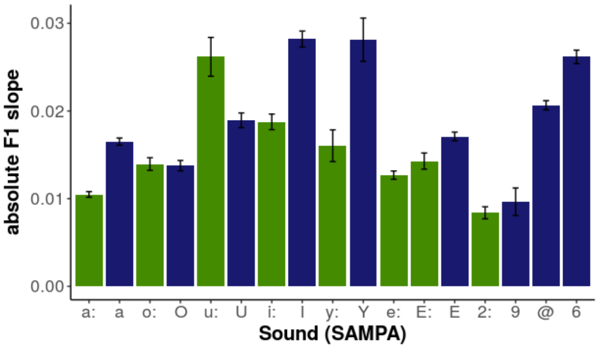

Figure 3: *Absolute F1 slope per vowel identity (SAMPA). Lax vowels are displayed in blue and tense vowels in green.*

vowel duration ($r = 0.28$).

ID factors, such as SURPRISAL, and WORD FREQUENCY, as well as prosodic factors, i. e. STRESS, ARTICULATION RATE, and BOUNDARY were used as fixed effects. We included vocalic phonemic STATUS (tense or lax) and average vowel DURATION as additional control factors, which was calculated per vowel identity and log-transformed. Random factors were SPEAKER, WORD, PRECEDING CONTEXT and FOLLOWING CONTEXT. Here, context was defined as place of articulation. For consonants we used the levels coronal, dorsal, and labial. If context consisted of pause or vocalic context, this was also marked. All categorical variables were treatment-coded before they were entered into the model. Backward model selection procedure with maximal random structure was used to identify the largest converging model. During model selection random slopes were excluded from the LMM because of correlations with random intercepts or convergence errors.

For both F1 slope and VL, we found the same effects for the ID variables, global ARTICULATION RATE, and PHONEMIC STATUS. Easily predictable, high-frequency vowels showed less formant change than low-frequency words that were difficult to predict. For F2 slope, on the other hand, we found the same effect of WORD FREQUENCY, but not for SURPRISAL. Somewhat unexpectedly, there was a significant negative effect of biphone surprisal of the preceding context on F2 slope, although the relationship was positive in simple Pearson's $r$ correlation analysis ($r = 0.02$).

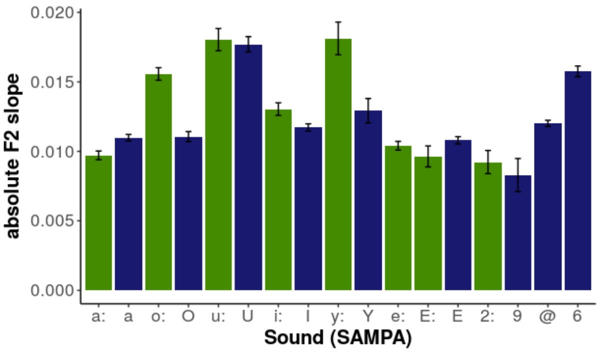

Figure 4: *Absolute F2 slope per vowel identity (SAMPA). Lax vowels are displayed in blue and tense vowels in green.*

As global tempo increased, the overall formant change in VL and F1 slope decreased. There was no effect of ARTICULATION RATE on F2 slope. Local ARTICULATION RATE was not significant in any of the models for VISC measures. Primary lexical STRESS had a significant positive effect on VL and F2 slope, but did not have an effect on F1 slope. While BOUNDARY had no effect on VL or F1 slope, we found that vowels showed more formant change in F2 when preceding a word boundary.

On average, lax vowels showed more formant change in VL and F1 slope than tense vowels. For F2 slope, this effect was opposite with tense vowels showing more spectral change. In the VL model, formant change was positively related to vowel DURATION: Longer vowels had longer formant vectors than shorter vowels. However, for F1 and F2 slope we found the opposite effect: DURATION was negatively related to absolute formant slopes.

The LMM with the largest effect size according to conditional pseudo $R^2$ was the F1 slope model with a total of 18.43 % explained variance. 11 % of that variance was explained by the fixed effects. Average vowel DURATION was the strongest predictor for F1 slope ($Var = 7.98\,\%$). SURPRISAL added 0.59 % and WORD FREQUENCY 0.11 % to the explained data variance in that model. Phonemic STATUS of the vowel had a similar effect size as WORD FREQUENCY ($Var = 0.10\,\%$), while global ARTICULATION RATE contributed only marginally to the model ($Var = 0.01\,\%$).

The LMMs for F2 slope and VL had a strong random structure, while the fixed effects only explained a small amount of variance in the data. Conditional pseudo $R^2$ added up to 12.59 % explained variance for VL and 17.48 % for F2 slope. Only 1.33 % of the data variance was accounted for by fixed effects in the F2 slope model, and even less in the VL model ($Var = 0.41\,\%$). For F2 slope, DURATION was the strongest predictor, followed by SURPRISAL ($Var = 0.08\,\%$), and WORD FREQUENCY ($Var = 0.03\,\%$). The prosodic factors primary lexical STRESS ($Var = 0.02\,\%$) and BOUNDARY were less effective ($Var = 0.004\,\%$). Phonemic STATUS did not add much to the model either ($Var = 0.002\,\%$), in contrast to the LMM for VL. Here, phonemic STATUS was the strongest fac-

Table 1: *Regression coefficients (Coeff.) and standard error (SE) of VISC LMMs. Only significant effects are reported.*

| VISC | Terms | Coeff. (SE) | | t-value |
|---|---|---|---|---|
| **VL** | Surprisal | $0.03$ | $(0.007)$ | $4.50^{***}$ |
| | Frequency | $-0.007$ | $(0.002)$ | $-3.85^{***}$ |
| | Stress | $0.02$ | $(0.008)$ | $2.86^{**}$ |
| | Global rate | $-0.01$ | $(0.004)$ | $-2.76^{**}$ |
| | Duration | $0.05$ | $(0.009)$ | $5.76^{***}$ |
| | Status | $-0.07$ | $(0.008)$ | $-8.54^{***}$ |
| **F1 Slope** | Surprisal | $0.003$ | $(0.001)$ | $13.55^{***}$ |
| | Frequency | $-0.0001$ | $(0.0001)$ | $-2.09^{*}$ |
| | Global rate | $-0.0003$ | $(0.0001$ | $-2.19^{*}$ |
| | Duration | $-0.02$ | $(0.0003)$ | $-61.98^{***}$ |
| | Status | $-0.001$ | $(0.0003)$ | $-3.70^{***}$ |
| **F2 Slope** | Surprisal | $-0.001$ | $(0.0001)$ | $-5.42^{***}$ |
| | Frequency | $-0.0001$ | $(0.0003)$ | $-4.39^{***}$ |
| | Stress | $0.001$ | $(0.0001)$ | $10.36^{***}$ |
| | Boundary | $0.001$ | $(0.0002)$ | $5.14^{***}$ |
| | Duration | $-0.004$ | $(0.0002)$ | $-21.68^{***}$ |
| | Status | $0.0006$ | $(0.0001)$ | $4.18^{***}$ |

tor with 0.21 % explained variance. WORD FREQUENCY and SURPRISAL both explained a total of 0.08 % variance, while the prosodic factors STRESS and global ARTICULATION RATE only summed up to 0.05 %. Average vowel DURATION also added to explained variance in VL ($Var = 0.03\,\%$)

With regard to additional LMMs testing larger n-phone dependencies, we tested the effect of triphone surprisal of the preceding context on VL. Surprisal was not significant in that model but showed a tendency for the same effect found in the model with biphone surprisal ($\beta = 0.008, SE = 0.004, t(39960) = 1.82, p = 0.07$). Neither absolute F1 nor F2 slope correlated positively with any of the other surprisal measures tested here. Therefore, no additional LMMs were calculated.

Interaction models were run for VL and F1 slope. We entered interaction terms for surprisal and global articulation rate in both models. According to ANOVA tests using the log likelihood output both models did not perform better than the baseline model without the interaction term. In the VL model, we ran a separate analysis including an interaction term for surprisal and stress. The interaction model performed better than the baseline model ($\chi^2(1) = 43.97, p < 0.001$). For vowels in stressed syllables, VL increased with higher surprisal values. The interaction of surprisal and stress added 0.13 % explained variance to the fixed effects of the VL model.

## 4. Discussion

This study investigated the impact of prosodic structure and information density on dynamic formant trajectories in German read speech. We found that word frequency was predictive of formant change in F1, F2, and VL. Vowels in high-frequency words showed less formant change than in low frequency words. Biphone surprisal of the preceding context was predictive of VL and F1 slope which was mirrored in the correlation analysis. The change in sign for surprisal in the F2 slope model despite a positive correlation ($r = 0.02$) may be explained by the complexity of the LMM and the weakness of the correlation. These findings showed that vowel formant change was significantly influenced by ID factors on phoneme and word level. We confirmed our hypothesis that vowels showed less formant change when they were easily predictable and occurred in high-frequent words.

Regarding the prosodic factors we observed a consistent effect of global articulation rate on VL and F1 slope, and of stress on VL and F2 slope. At accelerated speech rate there was less vocalic formant change than at slow speech rate. At a local level differences in speech rate did not explain variability in formant movement which was probably due to the collinearity between average vowel duration and local articulation rate. Vowels in syllables carrying primary lexical stress showed more formant change than in unstressed syllables. This was expected considering that vowels increase in duration and in their vowel dispersion when they stand in stressed lexical position [6]. There was an increase in F2 movement at the word boundary. This finding mirrors expansion of articulatory gestures in segments that undergo final-lengthening before a prosodic boundary [27].

F1 and F2 slope are measures of formant change relative to the duration of the vowel. Since the measure already includes vowel duration we found a negative relationship between absolute formant slope and average vowel duration. For instance, keeping the amount of change constant but doubling the duration of the vowel in which the change occurs leads to the relative amount of change being halved. VL, on the other hand, is not measured relative to the duration of the vowel, which is why we found a positive effect of vowel duration on VL, which in turn is in line with previous findings on VISC and duration variability [28].

We decided to include the phonemic status of the vowel in the LMM rather than vowel identity because of data sparsity and convergence issues of the model. However, we did not find consistent results for vowel tenseness across the VISC measures. Following [15] we expected to find more formant change in tense vowels and could confirm this hypothesis for F2 slope, but not for F1 slope and VL. In a more detailed post-hoc analysis of the German vowels with tense and lax pairs it became evident that differences in VL and F1 slope between these pairs were dependent on vowel identity (Figures 3 and 4). The binary coding of vowel identities into tense and lax vowels did not reveal those fine-grained differences. Also, these dynamic cues, which were apparently specific to vowel identity, helped listeners to identify tense and lax vowels in addition to information about inherent vowel duration [16].

## 5. Conclusions

Overall, formant change was equally affected by ID and prosodic structure. We only found an improvement in model performance when including an interaction between primary lexical stress and surprisal in the VL model. Here, vowels in stressed syllables showed an increase in VL when they were difficult to predict. However, the effect size of the fixed structure was low compared to the random structure. Phonological context, word and speaker identity were more informative in explaining variability in formant trajectories. This finding was expected considering the well-known effects of phonological context on formant movements.

## 6. Acknowledgements

This research was funded by the German Research Foundation (DFG) as part of SFB 1102 ‘Information Density and Linguistic Encoding’at Saarland University.

# 7. References

[1] M. Aylett and A. Turk, "Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei," *Journal of the Acoustical Society of America*, vol. 119, pp. 3048–3058, 2006.

[2] R. J. J. H. van Son and L. C. W. Pols, "An acoustic model of communicative efficiency in consonants and vowels taking into account context distinctiveness," in *Proceedings of ICPhS*, 2003, pp. 2141–2144.

[3] 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*, vol. 47, no. 1, pp. 31–56, 2004.

[4] R. S. Burdin and C. G. Clopper, "Phonetic reduction, vowel duration, and prosodic structure," in *Proceedings of ICPhS*, 2015.

[5] S. Kakouros and O. Räsänen, "Perception of sentence stress in speech correlates with the temporal unpredictability of prosodic features," *Cognitive Science*, vol. 40, no. 7, pp. 1739–1774, 2016.

[6] Z. Malisz, E. Brandt, B. Möbius, Y. Oh, and B. Andreeva, "Dimensions of segmental variability: interaction of prosody and information density in six languages," 2017, (in review).

[7] R. A. Fox and E. Jacewicz, "Cross-dialectal variation in formant dynamics of American English vowels." *The Journal of the Acoustical Society of America*, vol. 126, no. 5, pp. 2603–18, 2009.

[8] T. M. Nearey and P. F. Assmann, "Modeling the role of inherent spectral change in vowel identification," *The Journal of the Acoustical Society of America*, vol. 80, no. 5, pp. 1297–1308, 1986.

[9] S. A. Zahorian and A. J. Jagharghi, "Spectral shape features versus formants as acoustic correlates for vowels," *Journal of Acoustical Society of America*, vol. 94, no. 4, pp. 1966–1982, 1993.

[10] R. A. Fox, E. Jacewicz, and J. Hart, "Pitch pattern variations in three regional varieties of American English," in *Proceedings of INTERSPEECH*, no. August, 2013, pp. 123–127.

[11] H. Schoormann, W. Heeringa, and J. Peters, "Regional variation of Saterland Frisian vowels," in *Proceedings of ICPhS*, no. 1, 2015, pp. 1–5.

[12] R. Mayr and H. Davies, "A cross-dialectal acoustic study of the monophthongs and diphthongs of Welsh," *Journal of the International Phonetic Association*, vol. 41, no. 1, p. 1–25, 2011.

[13] T. L. Johnson and W. Strange, "Perceptual constancy of vowels in rapid speech," *The Journal of the Acoustical Society of America*, vol. 72, no. 6, pp. 1761–1770, 1982.

[14] J. W. Stack, W. Strange, J. J. Jenkins, W. D. Clarke, and S. A. Trent, "Perceptual invariance of coarticulated vowels over variations in speaking rate," *The Journal of the Acoustical Society of America*, vol. 119, no. 4, pp. 2394–2405, 2006.

[15] I. Lehiste and G. E. Peterson, "Transitions, glides, and diphthongs," *Journal of Acoustical Society of America*, vol. 33, no. 3, pp. 268–277, 1961.

[16] W. Strange and O.-S. Bohn, "Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies," *The Journal of the Acoustical Society of America*, vol. 104, no. 1, pp. 488–504, 1998.

[17] F. Schiel. (1997) Siemens Synthesis Corpus - SI1000P. [Online]. Available: https://www.phonetik.uni-muenchen.de/Bas/BasSI1000Peng.html

[18] T. Kisler, R. U. D., F. Schiel, C. Draxler, B. Jackl, and N. Pörner, "BAS Speech Science Web Services - an update of current developments," in *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, Portoroz, Slovenia, 2016.

[19] M. Baroni and A. Kilgarriff, "Large linguistically-processed web corpora for multiple languages," in *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics*, ser. EACL '06. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006, pp. 87–90.

[20] G. Möhler, A. Schweitzer, M. Breitenbücher, and M. Barbisch. (2000) IMS German Festival (Version: 1.2-os). University of Stuttgart: Institut für maschinelle Sprachverarbeitung (IMS).

[21] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 2017. [Online]. Available: http://www.fon.hum.uva.nl/praat

[22] J. Harrington, *The Phonetic Analysis of Speech Corpora*. Wiley-Blackwell, 2010.

[23] P. Adank, R. Smits, and R. van Hout, "A comparison of vowel normalization procedures for language variation research," *Journal of the Acoustical Society of America*, vol. 116, no. 5, pp. 3099–3107, 2004.

[24] A. Stolcke, "SRILM - an extensible language modeling toolkit," in *Proceedings of the International Conference on Spoken Language Processing*, vol. 2, Denver, 2002, pp. 901–904.

[25] A. Bell, J. 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, no. 1, pp. 92–111, 2009.

[26] 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.

[27] D. Byrd, "Articulatory vowel lengthening and coordination at phrasal junctures," *Phonetica*, vol. 57, no. 1, pp. 3–16, 2000.

[28] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels," *The Journal of the Acoustical Society of America*, vol. 97, no. May, pp. 3099–3111, 1995.