Speakers coarticulate less when facing real and imagined communicative difficulties: An analysis of read and spontaneous speech from the LUCID corpus

Zhe-chen Guo[1], *Rajka Smiljanic*[1]

[1]Department of Linguistics, The University of Texas at Austin
zcadamguo@utexas.edu, rajka@austin.utexas.edu

Abstract

This study investigated coarticulation of read and spontaneous speech in different communicative contexts from the LUCID corpus. Spontaneous speech samples were from Southern British English speakers who completed an interactive spot-the-differences task with no communicative barrier (NB), with their voice vocoded (VOC), and with a partner who heard their speech in babble (BABBLE) or was a non-native English speaker (L2). The same speakers also read sentences in a casual (READ-CO) and clear (READ-CL) speaking style. Tokens of a pre-defined set of keywords were extracted from the speech samples and consonant-vowel sequences in these tokens were analyzed using a whole-spectrum measure of coarticulation. Results showed that coarticulatory resistance in the six communicative contexts from highest to lowest was: BABBLE > VOC, L2, READ-CL > NB, READ-CO. Thus, in response to communicative barriers, be they real or imaginary, speakers coarticulated less, in line with the models of targeted speaker adaptations (the H&H theory [1] and Adaptive Speaker Framework [2]).

Index Terms: coarticulation, speaking style, spectral distance, spontaneous speech corpus

1. Introduction

Speakers continuously vary their speech output in response to the communicative context, reflecting a dynamic balance between hypospeech and hyperspeech (the H&H theory [1]; Adaptive Speaker Framework [2]). When comprehension is diminished, the speaker spontaneously shifts away from hypospeech and produces hyperarticulated listener-oriented clear speech forms aimed at addressing specific perceptual challenges faced by their interlocutor. Relative to casual or conversational speech, clear speech includes several intelligibility-enhancing acoustic-phonetic modifications, such as longer duration, vowel space expansion, and targeted segment hyperarticulations in response to phonological confusion (see [3]–[5] for reviews). Yet, with regards to how coarticulation, or the gestural overlap between neighboring sounds, varies with speaking style changes, acoustic analyses are sparse and yield mixed results. The main goal of this study is to examine coarticulation of read and spontaneous speech elicited in response to different communicative barriers using a whole-spectrum measure of coarticulation [6], [7].

In the H&H framework, coarticulation is a low-cost motor behavior [8]. More effortful listener-oriented speaking styles are thus expected to show less coarticulation compared to casual speech. For example, Moon and Lindblom [9] found larger F_2 displacements for hyper-articulated vowels compared to the vowels produced without attention to clarity in the /w_l/ frame, indicating less coarticulatory influence by the surrounding consonants. Similar reduction in coarticulation for vowels in syllables read in citation form compared to those that occurred spontaneously have been reported for Swedish, French, Spanish, and Catalan [10]–[12]. In contrast, Bradlow [13] found that coarticulation in English and Spanish clearly produced CV syllables was neither reduced nor exaggerated compared to conversationally produced syllables. Likewise, Matthies et al. [14] reported no effect of speaking style on CV coarticulation.

In all of the above studies, hyper-articulated speech was elicited in the absence of a true communicative intent. That is, the speakers followed instructions and spoke to an imagined or implied listener with perceptual difficulty. It is thus possible that the coarticulation findings were mixed since the speaking style adjustments were not explicitly motivated. More recently, Scarborough and Zellou [15] investigated vowel-nasal coarticulation in a "real listener" condition in which speakers completed an interactive task with a partner. They found that spontaneous clear speech in the real-listener condition showed more expanded vowel space and *greater* nasal coarticulation compared to read clear speech elicited with instructions to speak clearly.

The current understanding of how speaking style and coarticulation are related is still limited. Communicative intent introduced via the presence of a listener, as was done in Scarborough and Zellou's interactive task, does not necessarily imply a communicative barrier, such as competing speech or environmental noise which would induce intelligibility-enhancing speaking style adjustments. Here, we test how coarticulation is affected by speaking style modifications when speakers are experiencing communicative barriers. We analyzed recordings from the LUCID corpus [16]. The corpus consists of read conversational and clear speech elicited without a communicative partner and of spontaneous speech from talkers undertaking an interactive task in the presence of different communicative barriers (for more detail see Section 2).

To quantify coarticulation, we applied a whole-spectrum measure introduced by Gerosa et al. [6] and validated by Cychosz et al. [7]. This measure differs from the measures used previously, such as F_2 changes or the A1-P0 value [17] used for vowel-nasal coarticulation, in that it captures a distance between two overall spectral shapes. It does not rely on formant tracking, which may be unreliable, and is better suited for a broad range of consonant manners. The analysis was performed on CV syllables in a pre-defined set of words. Based on the H&H model and several previous studies (e.g., [9], [12]), we predicted that hyper-articulated intelligibility-enhancing speech produced in response to the imagined perceptual difficulty on the part of the listener and in response to the real communicative barriers will show more coarticulatory resistance compared to speech produced in the absence of such difficulty or barriers.

2. The LUCID corpus

The LUCID corpus features spontaneous speech from 40 Southern British English speakers working in pairs to complete the Diapix task, an interactive spot-the-differences task. Picture pairs of three different themes, each containing 12 differences, were used to elicit productions of 12 English monosyllabic CV(C) keywords (e.g., *sea*, *beach*) for a total of 36 keywords. The keywords, based on which the picture pairs were designed, began with a stop (/p/ or /b/) or a fricative (/s/ or /ʃ/) and consisted of (near) minimal word pairs (e.g., *sign*/*shine*, *pear*/*bear*). The 40 speakers participated in five recording sessions on separate days. On the first day, each participant pair completed the task in good listening conditions with no communicative barrier (NB). On the second day, they completed the task, first with one participant's speech vocoded and then with the other participant's speech vocoded (VOC). On the third day, half of the participants were paired with a native English-speaking experiment confederate who heard their speech in eight-talker babble (BABBLE). The other half worked with an experiment confederate who was a non-native speaker of English (L2).

In the last two sessions, the participants read keyword-bearing sentences (e.g., *The young children loved the* <u>*beach*</u>) individually without an actual interlocutor. They were instructed to read "casually as if talking to a friend" (READ-CO) in the first of these sessions and "clearly as if talking to someone who is hearing impaired" (READ-CL) in the second one.

3. Coarticulation analysis

3.1. Preprocessing

The LUCID recordings, along with their word-aligned Textgrids, were retrieved from SpeechBox [18] and had a total duration of 49 hours. Recordings from all the Diapix sessions except for the L2 one consisted of two channels, one for each talker in each pair. The two channels were extracted and saved as separate files. For the NB condition, the resulting single-channel files for the two talkers from the same original recording were both included in the analysis. For the VOC condition, the audio files of each speaker when their speech was vocoded were included. Similarly, for the BABBLE condition, only the files of the speakers whose speech was presented in talker babble to the listeners were included. No channel segregation was done for the other conditions: the original audio files from the two read-speech conditions contained a single channel, so did those from the L2 condition, in which the non-native confederates were not recorded. The recordings of eight speakers under the READ-CL condition were not available on SpeechBox and those of two speakers under this condition were excluded as the last few words in several sentences were clipped.

To obtain phone-aligned annotations, which were required by the spectral distance analysis described below, the Prosodylab-Aligner [19] was used to force-align each segment. The recordings were downsampled from the native sampling rate of 44.1 kHz to 16 kHz to facilitate the alignment and the outputs were word- and phone-aligned TextGrids. Only the 36 keywords were included in our analysis.

3.2. Spectral distance measurement

We followed the procedure for computing spectral distances used in [7]. First, each segment was partitioned into 25.6-ms frames with a step size of 10 ms. Short-time Fourier transform was applied to each frame to obtain a power spectrum, which was then passed through a Mel filter-bank and log-transformed. The values were averaged across all the frames in each segment to obtain an average log-Mel spectral vector representing the power distribution of that segment over different frequencies. The spectral distance between the consonant and vowel in a CV syllable (d_{cv}) was calculated as the Euclidean distance between the average log-Mel spectral vector of the consonant ($\bar{x}_c$) and that of the vowel ($\bar{x}_v$):

$$d_{cv} = \sqrt{(\bar{x}_c - \bar{x}_v)^2} \qquad (1)$$

This process was repeated for all CV syllables in all tokens of the 36 keywords. Stronger CV coarticulation should result in greater spectral similarity and hence a smaller d_{cv}. Coarticulatory resistance, on the other hand, should reduce spectral similarity and increase d_{cv}. The analysis was conducted on the recordings at the original sampling rate of 44.1 kHz with a custom Python script loading functions from the librosa package [20]. A total of 17,499 d_{cv} measurements were taken. Table 1 shows the numbers of measurements and talkers analyzed for each condition.

Table 1: *Numbers of talkers and spectral distance measurements by condition.*

Condition	No. of talkers	No. of measurements
NB	40	2,080
VOC	40	3,249
BABBLE	20	1,484
L2	20	1,857
READ-CO	40	5,096
READ-CL	30	3,733

4. Results

Shown in Figure 1 are the mean and individual d_{cv} values by condition. To address our questions, we fitted a Bayesian hierarchical generalized linear model with the brms package [21] of R [22]. A Gamma likelihood (with a log-link) was used in the model as distances are constrained to real non-negative values. The population-level effect of main interest was condition, dummy-coded with READ-CO as the baseline. In addition, since repeated or high-frequency words tend to be acoustically reduced [23]–[26], we also included, for each keyword, its repetition count (the *n*th time a keyword was produced by the same speaker in each picture task) and frequency (log frequency per million words according to the CLEARPOND database [27]) as population-level effects. Both these predictors were scaled. The group-level effects included intercepts and slopes for condition varying by participant as well as by CV sequence.

As there were no similar previous investigations, weakly informative regularizing priors were used on all parameters: Normal(0, 5) for the intercept, Normal(0, 1) for the other population-level effects, Gamma(0.01, 0.01) for the shape parameter, and $\text{Normal}_{+}(0, 1)$ for standard deviations of the group-level effects. Following [28], we also assigned a weakly informative regularizing prior (the LKJ-correlation prior [29]:

LKJ(2)) to the correlations between the group-level effects. The joint posterior distribution was approximated by Markov chain Monte Carlo samples obtained by four chains, each with 4,000 iterations and 1,000 warm-ups. The $\hat{R}$ statistic [30] was one for all parameters, indicating chain convergence.

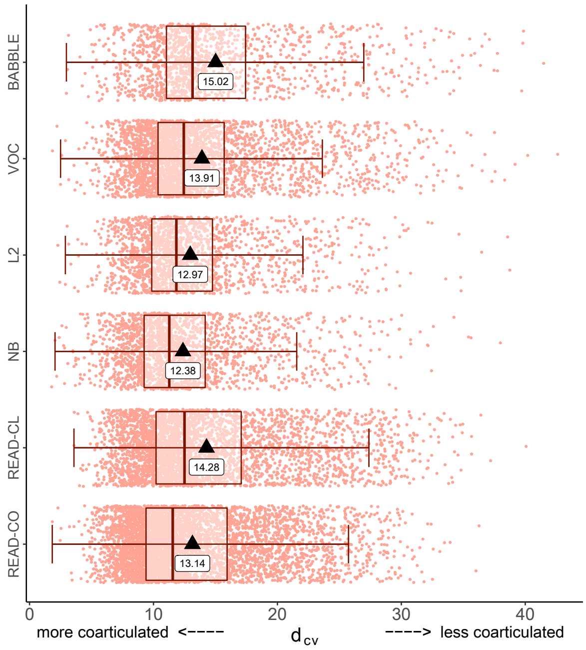

Figure 1: *CV coarticulation for each communicative condition: log-Mel spectral distance for each token (dots) along with the mean (triangle) and boxplot.*

Figure 2 gives the marginal posterior distributions of the population-level parameters. Inferences about the effect of a factor or difference between conditions were made by examining whether the 95% highest density interval (HDI) of the relevant posterior distribution excluded zero as a credible value. The marginal posterior on frequency fell almost entirely in the negative region (mean = −0.007, 95% HDI = [−0.011, −0.003]) and its 95% HDI excluded zero, indicating that CV syllables in repeated words tended to have a smaller d_{cv} and be more coarticulated. On the other hand, there was no evidence that word frequency affected coarticulation (mean = −0.009, 95% HDI = [−0.019, 0.001]).

The main question of the study was to examine whether within each task type (sentence reading or Diapix), coarticulatory resistance increased in response to communicative barriers, be they real or imaginary. To address this question, we computed the posterior distribution of mean d_{cv} for each condition. Two conditions were assumed to differ significantly in d_{cv} when the posterior distribution of their mean differences did not include zero in its 95% HDI. The posterior of mean differences between the READ-CL and READ-CO conditions (mean = 1.296; 95% HDI = [0.653, 1.930]) revealed that the former had a larger d_{cv}, suggesting that the speakers coarticulated less when instructed to speak clearly despite the absence of a real interlocutor. Similar coarticulatory resistance was found in the Diapix task: compared with those in the NB condition, d_{cv} values increased in the VOC (mean = 1.724; 95% HDI = [1.055, 2.407]), L2 (mean = 1.070; 95% HDI = [0.276, 1.860]), and BABBLE (mean = 2.823; 95% HDI = [1.634, 4.046]) conditions. BABBLE had a higher d_{cv} than VOC (mean = 1.100; 95% HDI = [0.155, 1.998]) and L2 (mean = 1.753; 95% HDI = [0.678, 2.903]), but the latter two did not differ (mean = 0.654; 95% HDI = [−0.101, 1.340]).

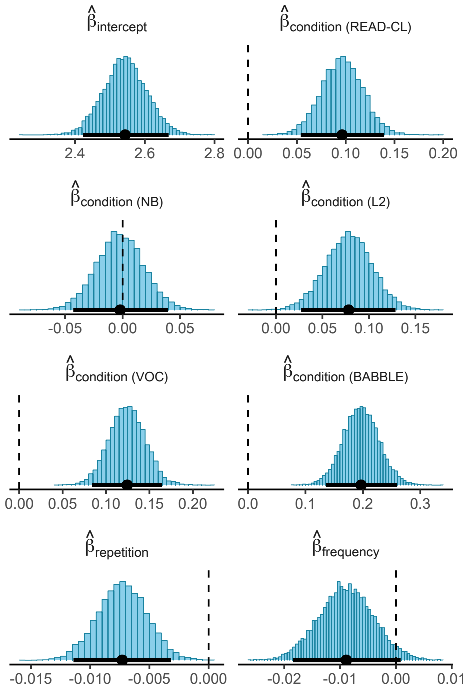

Figure 2: *Marginal posterior distributions of population-level parameters. Dots indicate their means and bars represent their 95% HDIs. Dashed line marks zero.*

To further explore differences between the conditions, we examined whether task type—or more essentially, a true communicative intent introduced via the presence of a partner—affected coarticulation when communicative barriers were either present or not. There was no evidence that NB and READ-CO differed in d_{cv} and hence in the degree of coarticulation (mean = −0.031; 95% HDI = [−0.586, 0.466]). Nor was there a difference between READ-CL on the one hand and L2 (mean = −0.256; 95% HDI = [−1.128, 0.632]) or VOC (mean = 0.397; 95% HDI = [−0.283, 1.074]) on the other hand. BABBLE had a higher d_{cv} than READ-CL (mean = 1.497; 95% HDI = [0.372, 2.669]), showing the least amount of coarticulation among all conditions.

Taken together, the findings suggested the following ranking of the conditions in terms of spectral distance across the two segments in CV syllables: BABBLE > VOC, L2, READ-CL > NB, READ-CO, where > means "has a higher d_{cv} (i.e., less coarticulation) than."

5. Discussion

By analyzing recordings from the LUCID corpus, this study explored the extent to which coarticulation varies as a function of different communicative contexts. In general, speakers coarticulated less in speech in response to both imagined perceptual difficulty on the part of the listener and real communicative barriers, whether or not a real listener was present. Such a finding is compatible with the H&H theory's view of coarticulation as a low-cost motor behavior and several empirical demonstrations (e.g., [9]–[12]) that coarticulation is reduced in more hyperarticulated speech. The relatively increased coarticulation (as reflected in lower d_{cv}) in the barrier-free conditions (i.e., READ-CO and NB) may be a consequence of shorter segment durations. Indeed, shorter segments in faster speech tend to be more coarticulated [31], and Hazan and Baker's [32] acoustic analysis of the LUCID corpus reveals that mean word durations in the barrier-free conditions are shorter than those in the barrier-present conditions. However, segment shortening does not have to be accompanied by increased coarticulation, which some speakers can avoid by speeding up their articulation [33]. This suggests some independence between shorter segment duration and increased gestural overlap, although, as the H&H model posits, the two tend to go hand in hand as the speaker shifts to hypospeech when comprehension is easy.

Additionally, the results revealed that the presence of a listener did not impact coarticulation across the barrier-free conditions: READ-CO and NB conditions were not significantly different. In contrast, different levels of coarticulatory resistance were found across the barrier-present conditions: BABBLE showed less coarticulation than VOC and L2, which were on par with READ-CL. The difference may reflect the variety in the type and extent of communicative challenges in realistic interactions with a listener, in response to which speaker flexibly adjust their speech patterns [2]. In line with this view is Hazan and Baker's [32] acoustic analysis of the BABBLE and VOC conditions. Although the two conditions did not differ significantly in mean word duration and F2 range, BABBLE was more exaggerated in all the other acoustic measures: it had greater median F0, F0 range, F1 range, and mean energy in the mid-frequency region. BABBLE speech may be produced with greater effort, consistent with its being less coarticulated than VOC. This demonstrates that talkers modify their speech in a selective manner to improve intelligibility in specific listening conditions.

Another issue that merits discussion concerns how the findings of this study relate to those from Scarborough and Zellou [15], who investigated anticipatory nasal coarticulation. In terms of experimental settings, their real listener condition, in which there was no communicative barrier, is more directly comparable to the NB condition from LUCID. Thus, insofar as coarticulation in the NB and READ-CL conditions is concerned, our results agree with theirs as we showed that spontaneous speech to a real-listener (in a barrier-free context) was more coarticulated compared to read speech elicited by instructions to speak clearly. If the coarticulatory patterns from the other conditions also generalize to anticipatory nasal coarticulation, then it would be expected that speakers may actually produce vowels with less anticipatory nasalization when talking to a listener but simultaneously experiencing communicative barriers.

Nonetheless, comparisons with Scarborough and Zellou's study are complicated by the vowel space results. While they reported more expanded vowel space in their real listener condition than in their read clear speech condition, Hazan and Baker [32] found the opposite: less expanded vowel space in the presence of a listener but with no barrier (NB) compared to READ-CL. It is unclear why changes in vowel space from a listener-present, barrier-free context to a context with an imagined interlocutor with a perceptual difficulty are in opposite directions across these studies. It is possible that the speech produced in Scarborough and Zellou's real listener condition was in fact relatively hyperarticulated and should be considered as comparable to that in conditions like BABBLE rather than NB. Yet, this raises another question as to why such hyperarticulated speech showed increased nasal coarticulation, contrary to the H&H theory and the current findings obtained with the whole-spectrum measure. At least for American English, English vowels are greatly nasalized by adjacent nasals relative to vowels in other languages (e.g., Thai) [34] and the extent of the nasalization varies substantially across English speakers [35]. It may be that the coarticulatory pattern of vowel-nasal sequences in hyperarticulated speech is different from that of other consonant manners. However, since the current data do not include nasal consonants, it still remains to be determined whether there is a difference in coarticulation between vowel-nasal sequences and other segment combinations in listener-oriented speech.

6. Conclusions

Using a whole-spectrum measure of coarticulation, an analysis of read and spontaneous speech from the LUCID corpus revealed the following ranking of communicative contexts in terms of their coarticulatory resistance (from highest to lowest): BABBLE > VOC, L2, READ-CL > NB, READ-CO. Speech in response to communicative barriers, whether they are real or not, shows increased coarticulatory resistance relative to speech in the absence of such barriers, in line with the H&H theory. The current findings also show that READ-CL is similar to the other real barrier conditions in showing reduced coarticulation. This suggests that READ-CL is a good proxy for investigating hyperarticulated listener-oriented speaking style when conducting more interactive tasks is not possible. The findings seem consistent with previous research on anticipatory nasal coarticulation, but the picture is complicated as the patterns of vowel space expansions are also considered. Further investigations are needed that include various segment types, systematically vary communicative contexts, and compare different measures of coarticulation. An open question is how decreased coarticulation, in addition to other clear speech modifications, contributes to segment identification and word recognition. Preliminary evidence suggests that reduced coarticulation in clear speech may contribute to better word segmentation in quiet but not in noise [36]. An improved understanding of coarticulatory patterns in listener-oriented clear speech would inform theories of phonetic variation and talker-listener adaptation. It would also have implications for implementing signal processing algorithms for different communicative situations in a more naturalistic manner.

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: Springer Netherlands, 1990, pp. 403–439.

[2] E. Buz, M. K. Tanenhaus, and T. F. Jaeger, "Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations," *Journal of Memory and Language*, vol. 89, pp. 68–86, 2016.

[3] M. K. Pichora-Fuller, H. Goy, and P. Van Lieshout, "Effect on speech intelligibility of changes in speech production influenced by instructions and communication environments," *Seminars in Hearing*, vol. 31, no. 2, pp. 77–94, 2010.

[4] R. Smiljanić and A. R. Bradlow, “Speaking and hearing clearly: Talker and listener factors in speaking style changes,” *Language and Linguistics Compass*, vol. 3, no. 1, pp. 236–264, 2009.
[5] R. M. Uchanski, “Clear speech,” in *The Handbook of Speech Perception*, D. Pisoni and R. Remez, Eds. Malden, MA: Blackwell, 2005, pp. 207–235.
[6] M. Gerosa, S. Lee, D. Giuliani, and S. Narayanan, “Analyzing children’s speech: An acoustic study of consonants and consonant-vowel transition,” in *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings*, 2006, pp. 393–396.
[7] M. Cychosz, J. R. Edwards, B. Munson, and K. Johnson, “Spectral and temporal measures of coarticulation in child speech,” *The Journal of the Acoustical Society of America*, vol. 146, no. 6, pp. EL516–EL522, 2019.
[8] E. Farnetani and D. Recasens, “Coarticulation and connected speech processes,” in *The Handbook of Phonetic Sciences: Second Edition*, W. J. Hardcastle, J. Laver, and F. E. Gibbon, Eds. Blackwell Publishing, 2010, pp. 316–352.
[9] S. J. Moon and B. Lindblom, “Interaction between duration, context, and speaking style in English stressed vowels,” *The Journal of the Acoustical Society of America*, vol. 96, no. 1, pp. 40–55, 1994.
[10] D. Duez, “Second formant locus-nucleus patterns: An investigation of spontaneouos French speech,” *Speech Communication*, vol. 11, no. 4–5, pp. 417–427, Oct. 1992.
[11] D. Poch-Olivé, N. Fernandez-Guitierrez, and G. Martinez-Dauden, “Some problems of coarticulation in CV stop syllables in Spanish and Catalan spontaneous speech,” in *Proceedings of Speech Resarch ‘89*, 1989, pp. 111–115.
[12] D. Krull, “Consonant-vowel coarticulation in spontaneous speech and in reference words,” *Speech Transmission Laboratory Quarterly Progress and Status Report*, vol. 30, no. 1, pp. 101–105, 1989.
[13] A. R. Bradlow, “Confluent talker- and listener-oriented forces in clear speech production,” in *Laboratory Phonology 7*, C. Gussenhoven and W. Natasha, Eds. Berlin & New York: Mouton de Gruyter, 2002, pp. 241–273.
[14] M. Matthies, P. Perrier, J. S. Perkell, and M. Zandipour, “Variation in anticipatory coarticulation with changes in clarity and rate,” *Journal of Speech, Language, and Hearing Research*, vol. 44, no. 2, pp. 340–353, Apr. 2001.
[15] R. Scarborough and G. Zellou, “Clarity in communication: ‘Clear’ speech authenticity and lexical neighborhood density effects in speech production and perception,” *The Journal of the Acoustical Society of America*, vol. 134, no. 5, pp. 3793–3807, 2013.
[16] R. Baker and V. Hazan, “DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs,” *Behavior Research Methods*, vol. 43, no. 3, pp. 761–770, 2011.
[17] M. Y. Chen, “Acoustic correlates of English and French nasalized vowels,” *The Journal of the Acoustical Society of America*, vol. 102, no. 4, pp. 2360–2370, 1997.
[18] A. R. Bradlow, “SpeechBox.” .
[19] K. Gorman, J. Howell, and M. Wagner, “Prosodylab-aligner: A tool for forced alignment of laboratory speech,” *Canadian Acoustics*, vol. 39, no. 3, pp. 192–193, 2011.
[20] B. McFee *et al.*, “Librosa - audio processing Python library,” *Proceedings of the 14th Python in Science Conference*, pp. 18–25, 2015, [Online]. Available: http://conference.scipy.org/proceedings/scipy2015/pdfs/brian_mcfee.pdf.
[21] P. C. Bürkner, “brms: An R package for Bayesian multilevel models using Stan,” *Journal of Statistical Software*, vol. 80, no. 1, 2017.
[22] 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/.
[23] C. A. Fowler and J. Housum, “Talkers’ signaling of ‘new’ and ‘old’ words in speech and listeners’ perception and use of the distinction,” *Journal of Memory and Language*, vol. 26, no. 5, pp. 489–504, 1987.
[24] 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, no. 1, pp. 31–56, 2004.
[25] 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, no. 1, pp. 92–111, 2009.
[26] T. Q. Lam and D. G. Watson, “Repetition reduction: Lexical repetition in the absence of referent repetition.,” *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 40, no. 3, pp. 829–843, 2014.
[27] V. Marian, J. Bartolotti, S. Chabal, and A. Shook, “CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities,” *PLoS One*, vol. 7, no. 8, p. e43230, 2012.
[28] S. Vasishth, B. Nicenboim, M. E. Beckman, F. Li, and E. J. Kong, “Bayesian data analysis in the phonetic sciences: A tutorial introduction,” *Journal of Phonetics*, vol. 71, pp. 147–161, 2018.
[29] D. Lewandowski, D. Kurowicka, and H. Joe, “Generating random correlation matrices based on vines and extended onion method,” *Journal of Multivariate Analysis*, vol. 100, no. 9, pp. 1989–2001, 2009.
[30] A. Gelman and J. Hill, *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press, 2006.
[31] A. Agwuele, H. M. Sussman, and B. Lindblom, “The effect of speaking rate on consonant vowel coarticulation,” *Phonetica*, vol. 65, no. 4, pp. 194–209, 2009.
[32] V. Hazan and R. Baker, “Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions,” *The Journal of the Acoustical Society of America*, vol. 130, no. 4, pp. 2139–2152, 2011.
[33] D. P. Kuehn and K. L. Moll, “A cineradiographic study of VC and CV articulatory velocities,” *Journal of Phonetics*, vol. 4, no. 4, pp. 303–320, 1976.
[34] P. S. Beddor and R. A. Krakow, “Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation,” *The Journal of the Acoustical Society of America*, vol. 106, no. 5, pp. 2868–2887, 1999.
[35] G. Zellou, “Individual differences in the production of nasal coarticulation and perceptual compensation,” *Journal of Phonetics*, vol. 61, pp. 13–29, 2017.
[36] Z. Guo and R. Smiljanić, “Speaking clearly improves speech segmentation by statistical learning under optimal listening conditions,” under review.