ISAPh 2022, 4th International Symposium on Applied Phonetics
14-16 September 2022, Lund, Sweden

The role of pause location in perceived fluency and proficiency in L2 Finnish

Heini Kallio, Mikko Kuronen, Liisa Koivusalo

University of Jyväskylä, Finland
heini.h.kallio@jyu.fi, mikko.j.kuronen@jyu.fi

Abstract

Fluency is a commonly used descriptor of second language (L2) speaking skills. Unplanned and too frequent pauses, hesitations, and repetitions disrupt the flow of speech and can cause temporal irregularities at all levels of speech hierarchy, from syllable rate to phrasing. However, most studies that attempt to quantify fluency disregard pause locations. The current study investigates, whether and which pause locations affect the perceived fluency and proficiency in L2 Finnish.

Several pause parameters were computed from spontaneous L2 Finnish speech samples. Pause locations were investigated with regards to clauses and phrases as well as incomplete words. The effect of pause locations to human assessments of fluency and proficiency was investigated using linear regression models.

The results suggest that silent and filled pauses within phrases and pauses after incomplete words significantly reduce the fluency of L2 Finnish speech, whereas pauses within clauses may even have a positive effect on fluency and proficiency. The results support the role of phrases over clauses as bases for prosodic grouping in spontaneous Finnish, but further research is needed with native speakers of Finnish. Furthermore, the results encourage to investigate the role of pause location parameters in comprehensive models for L2 speech fluency.

Keywords: L2 fluency and proficiency, pause location, automatic assessment

1. Introduction

Second or foreign language (L2) fluency has been widely studied from the perspective of temporal features related to speed and pausing, and common findings suggest that these phenomena are strongly associated with perceived fluency as well as proficiency in L2 [1, 2, 3, 4, 5, 6]. There are, however, somewhat conflicting results regarding different types of pause measures: for example, pause rate have been found both significant [1, 2] and non-significant [7] in predicting fluency ratings. Studies have similarly inconsistent findings on pause duration [1, 2, 7]. Results on the role of pausing to the perception of foreign accent have also varied: in [8], the longer the pauses the stronger the perceived foreign accent, while in [9] native speakers of Finnish were discovered to have longer pauses than L2 speakers. Few studies have attempted to explain these disagreeing results by focusing on *pause locations*, and the findings in these studies suggest that pause location is an important aspect in perceived L2 fluency [10, 11, 12]. To our knowledge, there are no earlier studies on the effect of pause locations on perceived proficiency in L2 speech.

Fluent speakers manage to maintain the continuity of their message while speaking, and pausing has an important role in this. Location of pauses is governed by psycholinguistic models of L2 speech production [13] – that is, pauses should precede grammatical constituents such as clauses. However, it has been reported that L2 speakers pause more often within utterances than L1 speakers [14]. Previous studies have also noted that fluent L2 speakers tend to pause at grammatical junctures (e.g., between clauses), whereas disfluent L2 speakers often pause within clauses or other syntactic constituents [15, 16]. The level of the speaker's language skills can affect both the ability to form clauses and the pausing patterns of the speaker.

Research on the effect of pausing on the perceived fluency and proficiency in L2 Finnish is scarce. In a recent study, we measured temporal fluency from spontaneous L2 Finnish speech and found that the rate of silent pauses (> 250 ms) and average duration of composite breaks (> 250 ms) significantly affect the perceived fluency and proficiency [6]. However, this was not the case for all L2 speakers: some speakers with similarly high pause rate vary notably in their fluency and proficiency ratings. In the current study, we scrutinize this observation by analyzing pause locations in relation to perceived fluency and proficiency. Differing from most previous studies that measure L2 speech fluency, the purpose of the current study is to investigate only the role of pause location, type, and rate to assessed fluency and proficiency, and we intentionally disregard other fluency measures such as speech and articulation rate.

10.21437/ISAPh.2022-5

Instead, we reflect the results concerning pause location to earlier studies where also speed measures were analyzed [6, 17].

Although most previous studies have investigated the effect of pause location to speech fluency with regards to the occurrence between or within clauses [18, 10], fluent prosodic phrasing does not necessarily follow the grammatical structures of a clause. We claim that pauses can be common or even relevant within a clause with several constituents, and that speakers with higher language proficiency produce longer and more complex clauses [19], which would increase the pause frequency within clauses. Therefore, in the current study we investigate pauses between and within grammatical clauses as well as between and within phrases. The motivation is in studying new measures for developing a more accurate automatic assessment of spontaneous L2 Finnish speech. This study is part of the DigiTala project that investigates and develops automatic tools for spoken L2 assessment and practicing purposes [20].

2. Material and methods

2.1. Speech data and human assessments

The speech data consists of spontaneous speech samples produced by 200 L2 Finnish speakers: 147 adult and 53 younger learners (high school students aged 15–21). The adult L2 data was provided by the National Certificates of Language Proficiency tests for Finnish as a second language [21] for the DigiTala project [20] and is described in more detail in a previous study [6]. The data consists of responses to narrative tasks, where the speakers were instructed to speak 1.5 minutes on a given topic. The data from high school students, in turn, was collected in the DigiTala project during 2021 and is described in detail in [17]. This data also consists of responses to a narrative task. In the task, the speakers were instructed to describe for one minute a place that is important for them. The tasks in both data sets had supportive questions for the speakers. Despite the instructive response times, the duration of the responses varied from 16 to 90 seconds.

Expert assessments were collected for both speech data using a 7-point holistic and five 3-4-point analytic rating scales developed for the purposes of the DigiTala project. For the current study, assessments of overall proficiency level (holistic scale) and fluency (analytic scale) were used.

The assessment data was collected separately for the adults' and high school students' speech samples, and the assessment processes are described in detail in [6] and [17]. In these studies, the inter-rater reliability was tested with intraclass correlation coefficient (ICC) using the *irr* package in R [22]. The average type ICC was selected as reliability measure, since it takes into account the scope of disagreement by comparing individual ratings of a sample to the mean rating of the sample. The selected ICC value was > 0.90 in both data sets, indicating excellent inter-rater reliability. We consider this as a justification to use average ratings of fluency and proficiency as the dependent variables in the current study.

2.2. Pause parameters

Previous studies have investigated the effect of pause location to speech fluency mainly with regards to their occurrence between or within clauses [10, 11, 12]. Here we investigate pauses *between clauses* (BC), *within clauses* (WC), *between phrases* (BP), and *within phrases* (WP). We define clause in Finnish as a constituent that links a predicate to a subject or object. Phrase, in turn, is defined as a word or group of words that act together as a grammatical unit, but do not necessarily include a predicate [23]. We compute pauses between and within *noun phrases*, *verb phrases*, and *adverbial phrases*, although the results for different phrase types are not separated in the analysis. In addition, pauses *within words*, or between an incomplete and a corrected word (WW), were measured. The effect of pause location to both perceived fluency and proficiency is investigated.

We used Praat [24] to annotate both *silent pauses* (SP) and *filled pauses* (FP) with respect to their location (BC, WC, BP, WP, or WW). The pause threshold was set to 250 ms that has been commonly used in previous research on speech fluency to define pauses and separate speech runs [25, 7, 1, 19, 6]. A pause were categorized as filled, when more than 50% of the pause duration included filler or hesitation sounds. Consequently, a pause was categorized as silent when more than 50% of the pause duration was silence. The annotated pause intervals were extracted using a Praat script and the computation of the pause parameters was done in R [26]. For pauses between clauses (BC) and phrases (BP), we created three measures: relative proportion of pause type in sample (ratio), number of pauses in sample (frequency), and average duration of pauses in sample (mean). For pauses within clauses, phrases, and words, we created the same three measures plus average frequency of pauses per phrase in sample (rate). The parameters are referred to using abbreviations of pause location, the type of pause, and the

type of measure (e.g., BC-SP ratio = ratio of silent pauses between clauses).

2.3. Statistical analysis

The effect of acoustic parameters on fluency and proficiency ratings was studied using multiple linear regression (MLR) models with average ratings as dependent variables and pause parameters as independent variables. From the computed pause measures we excluded absolute frequencies from the statistical analysis, since this measure is dependent on the number of clauses/phrases or overall length of a sample and not comparable between long and short speech samples.

The models were run separately for fluency ratings and proficiency ratings. The simplest models were derived using a feature selection method stepAIC (implemented in the R package MASS [27]) that selects the model with least information loss based on the Akaike Information Criterion (AIC).

3. Results

The contribution of pause parameters on the ratings of proficiency and fluency was studied using a stepwise linear regression model with average ratings as a dependent variable and pause parameters as predictor variables. The models were fitted separately for fluency and proficiency ratings. Table 1 lists all predictive variables tested and summarizes the results of the models with predictor t-values and respective significance levels based on p-values as well as the adjusted R^2 of the final models.

For both fluency and proficiency ratings the most significant predictor was the ratio of silent pauses within phrases (WP-SP ratio) with a t-value of -6.93 for fluency, t-value of -4.96 for proficiency, and $p < 0.001$, indicating that the larger the proportion of silent pauses within phrases, the lower the proficiency and fluency. The ratio of silent pauses between phrases (BP-SP ratio), rate of filled pauses within phrases (WP-FP rate), ratio of filled pauses between phrases (BP-FP ratio), and ratio of silent pauses after incomplete words (WW-SP ratio) also showed significant negative effects on both ratings.

Interestingly, the rate of silent pauses within clauses (WC-SP rate) and ratio of silent pauses between clauses (BC-SP ratio) had small yet significant positive effects on both ratings, indicating that the more often pauses occur within clauses and the larger the proportion of pauses between clauses, the better the proficiency and fluency. In addition, the ratio of filled pauses within clauses (WC-FP ratio) was selected to the final model for proficiency with a small positive effect, but the effect remained statistically insignificant.

Table 1: Summary of the linear regression models with predictor t-values and adjusted R^2s. Predictors with no values were excluded from the final models. p-values: 0.1–0.05', 0.05–0.01*, 0.01–0.001**, < 0.001***.

Predictor	Proficiency	Fluency
WC-SP ratio	-	-
WC-SP rate	2.15*	2.23*
mean WC-SP	-	-
BC-SP ratio	1.75'	2.23*
mean BC-SP	-	-
WC-FP ratio	1.43	-
WC-FP rate	-	-
mean WC-FP	-2.66**	-
BC-FP ratio	-	-
mean BC-FP	-	-
WP-SP ratio	-4.96***	-6.93***
WP-SP rate	-	-
mean WP-SP	-	-
BP-SP ratio	-4.33***	-5.33***
mean BP-SP	-	-
WP-FP ratio	-	-
WP-FP rate	-3.50***	-4.16***
mean WP-FP	-	-
BP-FP ratio	-2.09*	-3.38***
mean BP-FP	-	-
WW-SP ratio	-3.96***	-4.32***
WW-SP rate	-	-
mean WW-SP	-	-
WW-FP ratio	-	-
WW-FP rate	-	-
mean WW-FP	-	-
Model R^2 (Adjusted)	0.39	0.43

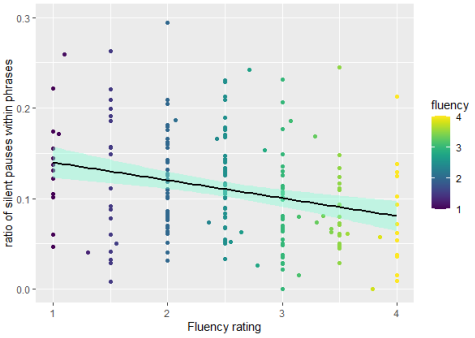

Figure 1: The linear trend between the ratio of silent pauses within phrases (WP-SP ratio) and average fluency ratings.

The final MLR models accounted for 43% of the

variance in fluency ratings (multiple R^2 = 0.45 and adjusted R^2 = 0.43) and 39% of the variance in proficiency ratings (multiple R^2 = 0.41 and adjusted R^2 = 0.39). The predictive power of single variables was investigated with simple linear models. Of the most significant predictors, WP-SP ratio explained 8% of fluency ratings and 5% of proficiency ratings, while BP-SP ratio accounted only for 3% of variance in fluency ratings and 1% of variance in proficiency ratings. WP-FP rate, in turn, explained 10% of fluency ratings and 12% of proficiency ratings. WW-SP ratio accounted for 9% of variance in fluency ratings and 5% of variance in proficiency ratings. The relations of WP-SP ratio, WP-FP rate, and WW-SP ratio to fluency ratings are shown as linear trends in Figures 1, 2, and 3. Interestingly, mean duration of WC-FPs account for 13% of the variation in proficiency ratings and 9% of the variation in fluency ratings, although the variable was not selected to the model predicting fluency.

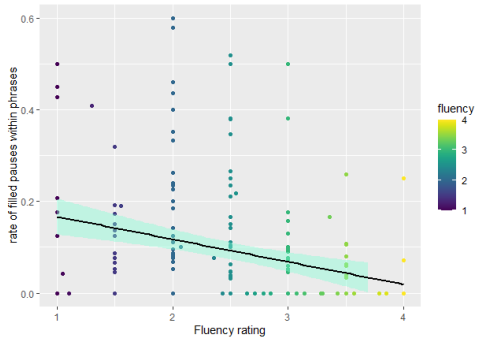

Figure 2: The linear trend between the rate of filled pauses within phrases (WP-FP rate) and average fluency ratings.

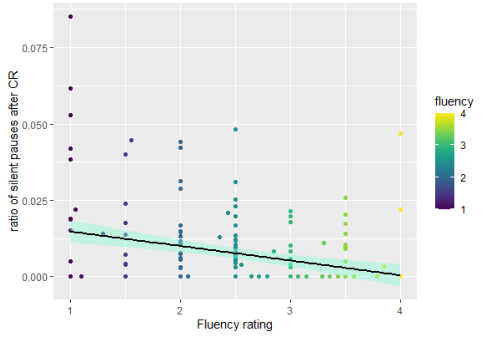

Figure 3: The linear trend between the ratio of silent pauses within words (WW-SP ratio) (or between an incomplete and corrected word) and average fluency ratings.

4. Discussion

This study investigated the role of pause locations in predicting perceived fluency and proficiency of L2 speakers of Finnish. To our knowledge, this study is the first to study pause locations in relation to overall oral proficiency. Moreover, there are only a handful earlier studies on the effect of pause location on fluency, most of them in L2 English. This study contributes to fluency research on scarcely studied L2, Finnish.

The effect of several pause parameters on fluency and proficiency ratings were investigated with multiple linear regressions. The MLR model for fluency ratings provided slightly higher explanatory power (adjusted R^2 = 0.43) than the one for proficiency (adjusted R^2 = 0.39) (cf. Table 1). This might be due to the more complex construct of proficiency compared to fluency, but also to the differences between the two scales: the fluency scale (1-4) was narrower than the proficiency scale (1-7), allowing less variation in the ratings and thus increasing the explanatory power of the fluency model.

Of the individual parameters, the ratio of silent pauses within phrases were the most significant type of disfluency in predicting both fluency and proficiency ratings. However, the parameter with the highest R^2 was the rate of filled pauses within phrase for fluency (WP-FP rate, explaining 10% of rating variation) and mean duration of filled pauses within clause for proficiency (mean WC-FP, explaining 13% of rating variation). This indicates that, alongside silent pauses within phrases, the use of filled pauses is a significant predictor of L2 Finnish fluency and proficiency, but the type of filled pause parameter depends on the assessed dimension. All these parameters had negative effects on ratings, meaning that the higher the value of the parameter, the lower the rating.

Interestingly, two clause-related parameters resulted in slightly positive effects on both fluency and proficiency: the rate of silent pauses within clauses and ratio of silent pauses between clauses. One possible explanation to this result is related to the duration of clauses: the longer the clause, the higher the probability of pausing becomes, and with longer utterances pausing is necessary for comprehensibility. Higher-proficiency speakers likely produce longer and more complex clauses than lower-proficiency speakers [19]. This can increase the number of pauses within clauses, but such pauses may be located between phrases or other syntactic constituents and hence do not impair with perceived fluency.

What connects the most significant parameters in the current study is their location with respect to phrases: silent or filled pauses occurring especially within phrases seem to be important in separating fluent speakers from disfluent. This supports

previous results on pause locations and L2 fluency [15, 16]. Pausing phenomena reflects the speaker's cognitive processing; pauses in the middle of syntactic constituents can be understood as difficulties in retrieving words or their spoken forms or difficulties in articulation. It is also noteworthy, that pause locations in terms of clauses did not provide similarly significant results than pause locations in terms of phrases. This result implies that, in Finnish, prosodic grouping rather follows the syntactic structure of phrases than clauses. However, native references were absent in this study, and thus we can not make reliable conclusions about the characteristic pausing patterns in L1 Finnish.

In previous studies pause locations have been mostly investigated in terms of clauses, but our result suggests that pause locations should be reconsidered in relation to phrases. However, our definition of phrase was sometimes at odds with the prosodic grouping in the current data: thinking pauses were relatively common after some conjunction words (such as "and") regardless of the fluency or proficiency level of the speaker. It should be further studied, whether such pauses are perceived as disfluencies or not.

The predictive power of the current model with pause parameters was not expected to be as high as of previous models with both speed and disfluency-related parameters. The comparison of the predictive power of MLR models done for the same speech data (Table 2) indeed suggests that speed-related fluency parameters are better estimates of perceived fluency than only pause-related parameters. However, the predictive power of the current model remained relatively high, which indicates that pause distributions by location could provide new information on L2 fluency. Moreover, the three studies referred to in Table 2 are not fully comparable, since the data of the two previous studies are combined in the current study. Whether pause distributions by location improve the predictive power of L2 fluency models, or whether the simplest measures such as speech-to-time ratio provide sufficient prediction accuracy should be further studied. If pause distributions by location increases the model accuracy, the next challenge is to automatically categorize pauses with respect to their syntactic location.

5. Conclusions

This study investigated, whether and which pause locations affect the perceived fluency and proficiency in L2 Finnish. Our results suggest that pause measures defined by their location with respect to *phrases* are more significant than pauses defined by their location with respect to *clauses* as predictors of L2 fluency as well as proficiency. Since clauses are more commonly used in previous studies on the effect of pause location on L2 fluency, the present study encourages to revise the methodology of the previous studies and consider phrases as the main syntactic constituents that guide prosodic grouping. In the future, we attempt to include pause location parameters to predictive models of L2 fluency and proficiency.

Table 2: Comparison of the MLR models predicting L2 Finnish fluency.

Study	**Predictors**	**Adj.** R^2
Current study (N = 200)	7 pause-location related measures	0.43
Kallio et al. 2022 (N = 147)	articulation rate, mean duration of utterance breaks, rate of SPs, duration of FPs, ratio of creak	0.46
Koivusalo 2022 (N = 53)	articulation rate, ratio of SPs	0.60

6. Acknowledgements

The authors would like to thank all the expert assessors for the ratings as well as the Finnish National Certificates of Language Proficiency for their help in enabling the use of part of the speech data and recruiting expert assessors. We also want to thank the Aalto University's DigiTala team for their help in creating the Moodle environment for collecting human ratings. The DigiTala project is funded by the Academy of Finland and the research consortium includes University of Helsinki (grant number 322619), Aalto University (grant number 322625), and University of Jyväskylä (grant number 322965).

7. References

[1] H. R. Bosker, A.-F. Pinget, H. Quené, T. Sanders, and N. H. De Jong, "What makes speech sound fluent? The contributions of pauses, speed and repairs," *Language Testing*, vol. 30, no. 2, pp. 159–175, 2013.

[2] C. Cucchiarini, H. Strik, and L. Boves, "Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech," *The Journal of the Acoustical Society of America*, vol. 111, no. 6, pp. 2862–2873, 2002.

[3] T. M. Derwing, M. J. Rossiter, M. J. Munro, and R. I. Thomson, "Second language fluency: Judgments on different tasks," *Language Learning*, vol. 54, no. 4, pp. 655–679, 2004.

[4] H. Kallio, J. Šimko, A. Huhta, R. Karhila, M. Vainio, E. Lindroos, R. Hildén, and M. Kurimo, "Towards the phonetic basis of spoken second language assessment: temporal features as indicators of perceived proficiency level," *AFinLA-e: Soveltavan kielitieteen tutkimuksia*, no. 10, pp. 193–213, 2017.

[5] H. Kallio, A. Suni, and J. Šimko, "Fluency-related temporal features and syllable prominence as prosodic proficiency predictors for learners of English with different language backgrounds," *Language and Speech*, vol. 65, no. 3, pp. 571–597, 2022.

[6] H. Kallio, R. Suviranta, M. Kuronen, and A. von Zansen, "Creaky voice and utterance fluency measures in predicting perceived fluency and oral proficiency of spontaneous L2 Finnish," *Speech Prosody: Proceedings of the 11th International Conference on Speech Prosody*, pp. 777–781, 2022.

[7] J. Kormos and M. Dénes, "Exploring measures and perceptions of fluency in the speech of second language learners," *System*, vol. 32, no. 2, pp. 145–164, 2004.

[8] P. Trofimovich and W. Baker, "Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech," *Studies in Second Language Acquisition*, vol. 28, no. 1, pp. 1–30, 2006.

[9] M. Toivola, "Vieraan aksentin arviointi ja mittaaminen suomessa [Assessment and measurement of foreign accent in Finnish]," *PhD Dissertation, University of Helsinki*, 2011.

[10] J. Kahng, "The effect of pause location on perceived fluency," *Applied Psycholinguistics*, vol. 39, no. 3, pp. 569–591, 2018.

[11] K. Saito, M. Ilkan, V. Magne, M. N. Tran, and S. Suzuki, "Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency," *Applied psycholinguistics*, vol. 39, no. 3, pp. 593–617, 2018.

[12] S. Suzuki and J. Kormos, "Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech," *Studies in Second Language Acquisition*, vol. 42, no. 1, pp. 143–167, 2020.

[13] J. Kormos, *Speech production and second language acquisition*. Routledge, 2014.

[14] N. H. De Jong, "Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency," *International Review of Applied Linguistics in Language Teaching*, vol. 54, no. 2, pp. 113–132, 2016.

[15] A. Riazantseva, "Second language proficiency and pausing: A study of Russian speakers of English," *Studies in Second Language Acquisition*, vol. 23, no. 4, pp. 497–526, 2001.

[16] J. Kahng, "Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall," *Language Learning*, vol. 64, no. 4, pp. 809–854, 2014.

[17] L. Koivusalo, "Phonetic Fluency in Finnish as a Second Language: Acoustic Analysis of High School Students' Spontaneous Speech," *Master's Thesis, University of Helsinki*, 2022.

[18] P. Tavakoli and P. Skehan, "Strategic planning, task structure and performance testing," in *Planning and Task Performance in a Second Language*, R. Ellis, Ed. John Benjamins, 2005, pp. 239–273.

[19] M. Kautonen and M. Kuronen, "Kvantitativa perspektiv på L2-tal på olika färdighetsnivåer [Quantitative perspectives on L2 speech on different skill levels]," *Folkmålsstudier*, vol. 59, pp. 11–40, 2021.

[20] M. Kautonen and A. von Zansen, "DigiTala research project: Automatic speech recognition in assessing L2 speaking," *Kieli, koulutus ja yhteiskunta*, vol. 11, no. 4, 2020.

[21] Finnish National Agency for Education. National Certificates of Language Proficiency. [Online]. Available: https://www.oph.fi/en/national-certificates-language-proficiency-yki

[22] M. Gamer, J. Lemon, M. M. Gamer, A. Robinson, and W. Kendall's, "Package 'irr'," *Various coefficients of interrater reliability and agreement*, vol. 22, 2012.

[23] A. Hakulinen, M. Vilkuna, R. Korhonen, V. Koivisto, T. R. Heinonen, and I. Alho, "VISK: Iso suomen kielioppi. Helsinki: Suomalaisen Kirjallisuuden Seura. Verkkoversio [The Great Grammar of Finnish. Online version]," *Online: http://scripta.kotus.fi/visk*, 2004.

[24] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer. [Computer program]. Version 6.0. 19," *Online: http://www. praat. org*, 2016.

[25] R. Towell, R. Hawkins, and N. Bazergui, "The development of fluency in advanced learners of French," *Applied Linguistics*, vol. 17, no. 1, pp. 84–119, 1996.

[26] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. [Online]. Available: https://www.R-project.org/

[27] B. Ripley, B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, D. Firth, and M. B. Ripley, "Package 'mass'," *Cran r*, vol. 538, pp. 113–120, 2013.