Speech Rhythm and Timing of a Speech-impaired Speaker

Tae-Jin Yoon[1], *Karin Humphreys*[2]

[1]Department of Linguistics and Languages, McMaster University, Canada
[2]Department of Psychology, Neuroscience, and Behaviour, McMaster University, Canada
tjyoon@mcmaster.ca, krh@mcmaster.ca

Abstract

We made quantitative rhythmic and timing measurements on speech samples obtained a 61 year-old monolingual female English speaker who is reported to have required a rare but possible case of Foreign Accent Syndrome (FAS). The phonetic characteristics of speech produced by individuals with FAS affects both suprasegmental and segmental properties. We used a large scale of speech samples to overcome difficulties of comparison due to the lack of speech samples before the accident. The results showed that the speaker has greater variability than control speakers, and that speech rate is slower, but that the rhythmic patterns were more to the stress timed. The results imply that the greater variability of the rhythmic and timing patterns in the speech-impaired speaker can be used as a means of identifying areas of speech production in the speech-impaired speaker.
Index Terms: disordered speech, Foreign Accent Syndrome (FAS), speech rhythm, tempo, rhythm metrics, speech rate

1. Introduction

We present detailed phonetic analyses of a 61 year-old monolingual female English speaker. She was born in St. Catherines, Ontario in Canada, and had resided in Nova Scotia since the fifth grade. She had lived in Eastern Atlantic Canada for most of her life, and was a retired special education teacher. The speaker acquired a speech disorder three years after suffering from sustained traumatic brain injury following a motor vehicle accident. Three years after the accident, her family members began observing noticeable changes in her speech. According to the family members, she woke up one morning speaking with slow and broken speech which sounded like a foreign accent. Her speech was regarded as often shifting from Atlantic Canadian English to Scottish English or as Southern US drawls at times [1]. It was usually the case with her, as well as other reported speakers with FAS, that the speech was fully intelligible and fluent but often resembled a 'foreign accent.' Thus, her disordered speech is regarded as a rare but possible case of Foreign Accent Syndrome (FAS).

Foreign accent syndrome (FAS) is a rare disorder characterized by the emergence of a new accent perceived as foreign by listeners. FAS is not due to the acquisition of a specific foreign accent, but to impairment of the segmental and suprasegmental linguistic abilities such as stress, rhythm, tempo, and vocal stress that make it possible to distinguish it from native language [2, 3]. Many studies have reported a slow rate of speech for individuals with FAS (see [3] for reference). These characteristics of slow speech may arise from long-term muscular adjustments of the vocal apparatus that lead to changes in articulatory, phonatory, and prosodic settings [4]. Given these long-term muscular adjustments, we may expect that the slow speech would be all-pervading in the speech of the impaired speaker. Unlike our expectation, however, considerable variability exists among reported cases of FAS [4, 5]. Also tt is typical for speakers with FAS that the articulatory and prosodic misproductions are sporadic and not consistent [4]. In many cases, our speaker demonstrated that accent changes were salient when she was extremely tired or anxious.

This observation raises a question of whether there is a way of reliably identifying regions of accent changes in a quantitative and objective manner. We hypothesize that the regions of accent changes can be partially but reliably identified by looking at the variability of the speech rhythm and tempo of the speech-disordered speaker. The defining characteristics of the FAS are segmental and prosodic deficits which result in the perception of a foreign accent by listeners. These abnormalities are related to the perception of rhythm patterns of the FAS speech. For example, many native English speakers with FAS have demonstrated rhythm changes suggesting a more syllabic timing pattern [3, 4]. The 'syllable-timed' pattern has been reported based on the observation that the speech samples exhibit unusually equal syllable durations, lack of reduction of unstressed vowels, and occurrence of epenthetic vowels, among others. Our observation of the speech samples in our speaker also indicates lack of reduction of unstressed vowels and the occurrence of epenthetic vowels. We also observed variability in gestural organization of the consonants, such that the intervocalic voiceless stops are often produced with longer closure duration and aspiration, word and utterance-final-stops are more than often fully released. Taken together, it is not clear what is the outcome of these interactions. Thus, a more reliable way of assessing the speaker's rhythm pattern is warranted. Furthermore the quantitative measures of the rhythm and timing patterns will reveal regions of broken speech in a reliable way.

We calculate speech tempo and rhythm, using measures such as speech rate (number of syllables per second), proportion of vocalic duration over the total duration of an utterance (%V), standard deviation of the vocalic and intervocalic consonantal durations (ΔV and ΔC, respectively), normalized and non-normalized pair-wise variability indices for vocalic intervals (nPVI-V) and consonantal intervals (rPVI-C), respectively, and ratio of the number of vocalic segments to the number of consonantal segments (V-to-C Ratio). A problem in dealing with speech-impaired speakers is that speech samples are only available after the speech-impairment, making it difficult to compare the characteristics of the speaker's speech patterns after the impairment and those before the impairment. In our study, we calculated the same measures made from 10 speakers from the Buckeye Corpus [6], and compared the measures between the speech-disordered speaker and 10 speakers in the Buckeye Corpus.

2. Analysis

2.1. Data

Upon request by the family members of the speech-impaired speaker, the second author visited the speaker in Halifax and collected the speech recordings through three one-hour sessions with the patient during the three day visit. The collected recordings range from simple read sentences to passages, and to spontaneous description of black and white line-art pictures. In this study, we report our analyses of the read sentences and passages with a special focus on speech rhythm and tempo. The recordings from spontaneous description of pictures are yet to be analyzed. All speech samples were audio-recorded in a quiet setting using a head-mounted unidirectional microphone and a professional mobile digital recorder.

The recordings were categorized into three speech styles. First, stimuli for segmental analyses were elicited, in which the speaker was reading a series of carrier phrases, such as the "I said ____ again" or "The next word is ____". This speech style is termed as **carrier phrase**. Second, stimuli for prosodic analysis included changing intonation elicitations (e.g. *We can all smile for the camera, can't we?"*) and sentences containing noun-verb stress-shifting word pairs (e.g., *The priest blessed the convert* vs. *They decided to convert the old building into an new school.*). This speech style is termed as **sentence**. Final stimuli at the sessions included standardized reading passages ('Comma Gets a Cure,' 'Arthur the Rat,' 'The Grandfather Passage,' and 'The Rainbow Passage' (e.g. *When the sunlight strikes raindrops in the air they act as a prism and form a rainbow* ....). We call these reading style **passage**.

All stimuli were presented in a random, mixed order in slideshow format using Microsoft PowerPoint. Once all the speech recordings and stimuli were collected, acoustic-phonetic features of the speaker's speech were documented, automatically segmented using a forced-alignment, and manually corrected. In all, 57 carrier phrases, 59 sentences, and 121 chunked files of the passages were used for the analysis. The duration of stimuli sums up to about 20 minutes of the non-silent portions of the recorded speech samples.

2.2. Buckeye Corpus as a reference

To compare the characteristics of the speaker's speech patterns after the impairment, the same measurements were made from ten speakers drawn from the Buckeye Corpus of conversational speech. The corpus contains high-quality recordings between 1999 and 2000 from speakers in Columbus, Ohio, conversing freely with an interviewer for approximately one hour. [6] The measures of 10 (7 female and 3 male) speakers serve as a reference to which the measures of the current study is compared. Even though the speakers in the Buckeye Corpus are native speakers of English from the mid-west, the speech style is not the same as that of impaired speakers. But in general, greater variability is observed in the spontaneous speech than in the reading style. If this tendency holds true in the current study, then we would expect more variability from samples in the Buckeye Corpus than from samples of the speech-impaired speaker. The duration of the 10 speakers (excluding the interviewer) ranges from 16 min to 40 min. The total duration of the 10 speakers is about 4.5 hours. Detailed analysis of the rhythm measures is found in [7].

2.3. Metrics for speech rhythm and tempo

Speech rhythm refers to the way languages are organized in time. In recent studies on speech rhythm, a series of acoustic metrics based on consonantal and vocalic duration have been designed to distinguish language according to putative stress and syllable-timed rhythmic categories. For example, Ramus and his colleagues [8] demonstrated that sentences in stress-timed (vs syllable-timed) languages had greater durational variability in "consonantal intervals" (sequences of abutting consonants regardless of syllable or word boundaries, as measured by $\Delta\mathbf{C}$ and a lower overall percentage of sentence duration devoted to vowels (**%V**). These differences likely reflect phonological factors such as the greater variety of syllable types and the greater degree of vowel reduction in stress-timed languages [8, 11]. In a similar vein, Pairwise Variability Index (PVI) measures the degree of durational contrast between successive elements in a sequence, and was also developed to explore rhythmic differences between "stress-timed" and "syllable-timed" languages [10]. Earlier studies have revealed that the normalized PVI of vowel durations (**nPVI-V**) and the raw PVI of consonantal durations in sentences is significantly higher in stress-timed languages (e.g. English) than in syllable-timed languages (e.g. French) [10]. The reason for this is, again, thought to be the greater degree of vowel reduction and the greater syllabic complexities around syllable onsets and codas in the stress-timed languages such as English.

Following the previous research tradition, we will adopt the metrics that measure the proportion of vocalic intervals (%V), the duration variability of vocalic intervals (e.g. ΔV, nPVI-V), duration variability of consonantal intervals (e.g. ΔC, rPVI-C), as well as speech rate for speech timing. Many of these metrics imply that the durational variability of vowels and consonants are greater for the stress-timed language than for the syllable-timed language. In order to verify whether this is really the case in our speech sample, we also calculate the ratio of the number of vocalic segments and the number of consonantal segments (V-to-C Ratio) to reflect the complexity of syllable structure. If the V-to-C Ratio is lower than 1, more consonantal segments are in the stimuli than vocalic segments (e.g. the V-to-C Ratio of CVCCCVC is 2/5=0.4). If the ratio is 1, equal number of consonantal and vocalic segments are in the stimuli.

In order to calculate these metrics of the speech-impaired speaker, we first segmented the speech files by applying an automatic phone-alignment to the speech samples. The obtained phone sequences are then manually corrected. The segmented information is used to calculate the V-to-C Ratio. Measurements of other metrics were made by collapsing a sequence of phone segments into vocalic and intervocalic consonantal intervals, where such intervals are defined as all consecutive segments of the same type irrespective of syllable or word boundaries. A few things to note are that utterance-medial pauses were excluded from rhythm measurement, but were included in the calculation of speech rate, and that phrase-final intervals were not excluded. The effect of final lengthening would result in greater vocalic variability.

3. Results

Table (1) presents the mean and standard deviation of the rhythm metrics for the speakers in the Buckeye Corpus and for the speech-impaired speaker in our study. In general, the speaker in our study has greater variability for all measures. It is also shown that the speech rate of the speech-impaired speaker

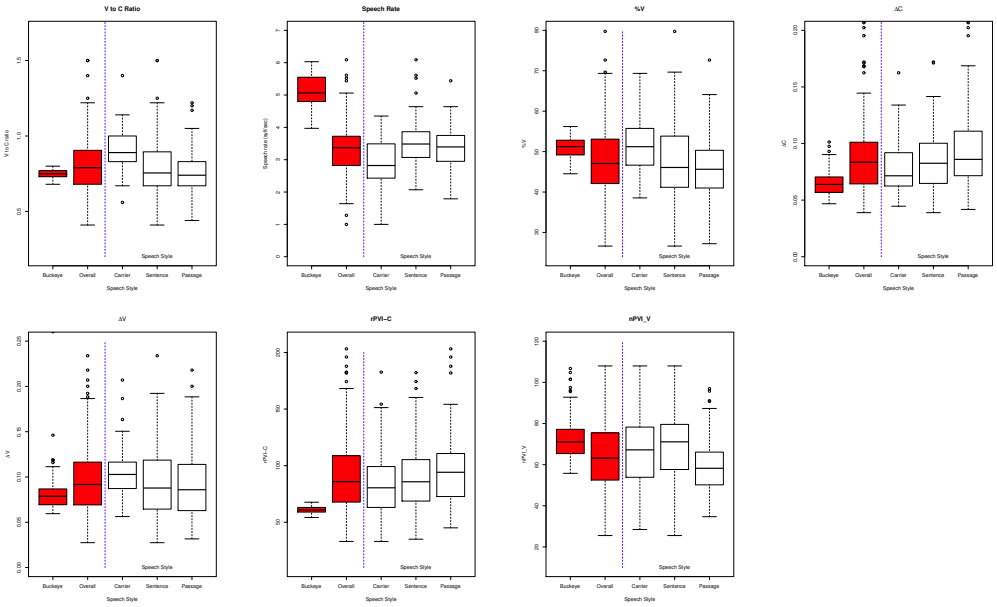


Figure 1: *The box-and-whisker plots of (1) V-to-C Ratio, (2) Speech Rate, (3) %V, (4) ΔC, (5) ΔV, (6) rPVI-C, and (7) nPVI-V. The leftmost two box plots are for the control speakers and the speech-impaired speaker. The three box plots on the right side of the dividing line show the central tendency and variability of the speech-disordered speaker categorized by speech styles of carrier phrases, sentences, and passages.*

is much lower than that obtained from the Buckeye Corpus. The slow speech rate is consistent with the previous studies [3].

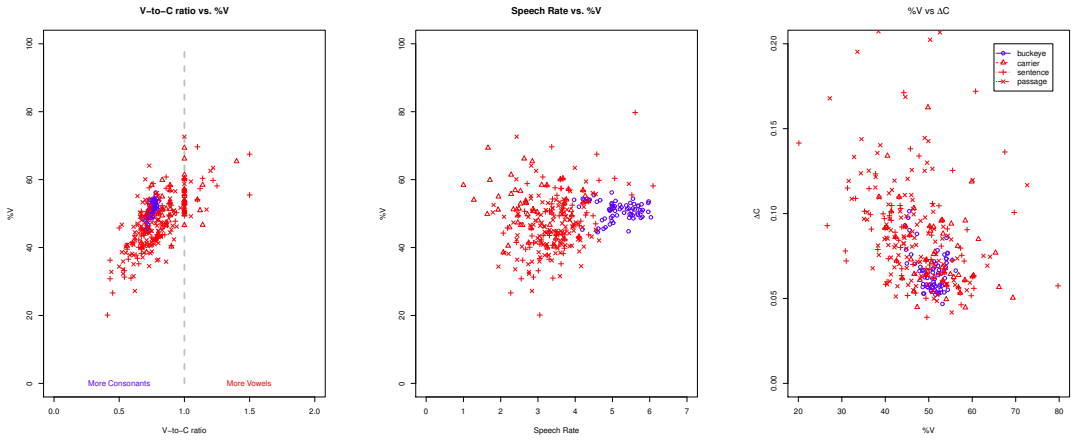
Table 1: *Mean and standard deviation (in the parentheses) for 10 speakers from the Buckeye Corpus and for the speaker in our current study.*

Rhythm Metrics	Buckeye Corpus	Speech-impaired speaker
Speech rate	5.12 (0.51)	3.31 (0.74)
V-to-C ratio	0.74 (0.02)	0.81 (0.20)
%V	50.85 (2.71)	47.63 (8.41)
ΔC	0.062 (0.011)	0.088 (0.034)
ΔV	0.084 (0.026)	0.099 (0.045)
rPVI-C	60.98 (2.88)	93.70 (37.80)
nPVI-V	73.49 (11.79)	63.70 (16.27)

The V-to-C Ratio being lower than 1 indicates that the structure of syllable (especially at the margins) in both data sets are comparably complex. But the %V and nPVI-V are lower and ΔC and rPVI-C are significantly higher for the disordered speaker than for the reference speakers in the Buckeye Corpus. The metric values more close to stress timed language together with slowness of speech rate indicates that the speech-impaired speaker is having more difficulty in coordinating consonantal gestures than the control speakers. While the variability of the vocalic segments in general (ΔV) is greater for the disordered speaker, the normalized variability of adjacent vocalic durations (nPVI-V) is lower than the speakers in the Buckeye Corpus. One of the factors that contribute to greater values of nPVI-V indicates greater variability in adjacent vowels, due to the stress-induced lengthening and reduction in unstressed syllable with a foot. Lower nPVI-V value for the speech-impaired speaker in Table (1) implies that less vowel reduction in unstressed syllable is observed for the speech-impaired speaker. These finding indicate that in general the speech-impaired speaker exhibits stress-timed rhythm pattern. This result is in opposition to the previous studies (e.g. [3]) reporting that the speakers of FAS in English is perceived to be more syllable-timed.

Figure (1) shows box plots for each rhythm metric. In each box plot, the leftmost plots show the central tendency and variability of measures between the Buckeye Corpus and the current study. Statistical analysis of Welch two sample t-test, with no assumption of homogeneity of variance, indicates that all measures result in significant differences between the speakers in the Buckeye Corpus and the speaker in our study. Even though it is not the main focus of the paper, the box plots for speech styles are also presented. That is, the three plots on the right side of the dividing line show the central tendency and variability of the speech-disordered speaker categorized by speech styles of carrier phrases, sentences, and passages.

As mentioned in the introduction, it is typical for speakers with FAS to exhibit sporadic articulatory and prosodic misproductions [4]. The speech-impaired speaker in our study also demonstrated changes occurs when she was extremely tired or anxious. These salient changes in her speech can be identified by looking at the relationship between rhythmic or timing metrics. For example, the relationship between V-to-C ratio and %V is shown in Figure (2a). The Buckeye Corpus has a value below 1 for its V-to-C ratio, which is expected for the canonical stress-timed language. As for the speaker of our current study, even though the mean of the V-to-C ratio is below 1, there are many instances in which the value goes over 1. This indicates that there are utterances containing a greater number of vocalic segments, resulting in simpler syllable structure. Some of these tokens are identified as having inserted vowels. For example, the CVC of 'bid' is produced with CVCV, where the final V is a short schwa-like vowel. As shown in Figure (2b), the %V



(a) V-to-C Ratio vs. %V (b) Speech Rate vs. %V (c) %V vs. ΔC

Figure 2: *Scatter plots for (a) V-to-C Ratio vs. Speech Rate, (b) Speech Rate vs. %V, and (c) %V vs. ΔC obtained from the Buckeye Corpus and the three distinct speech styles of the speech-impaired speaker.*

is not correlated highly with the speech rate, in line with the findings in Dellwo & Wagner (2003) [12]. Some tokens with low speech rate and high %V sounds rather syllable-timed than stress-timed. Figure (2c) illustrates the expected negative correlation between %V and ΔC. This scatter plot implies that as the value of %V gets smaller, the standard deviation of the consonantal intervals get larger. This is expected for the stress-timed language. But when we examined some of the tokens with very low %V and very high ΔC, we observed that the speaker had signs of difficulty in coordinating consonants. For example, in words like 'pretty,' the intervocalic consonant has very long closure duration followed by a long aspiration. Thus, we can use deviant values from the norm to identify the sporadic regions in which the speech-impaired speaker has difficulty in her production of spoken utterances.

4. Discussion and Conclusions

We have shown that the rhythm metrics indicates that the speaker in our study exhibits patterns that are aligned with more stress-timed rather than syllable-timed language. Further phonetic analyses were conducted to understand the impact of the speaker's speech characteristics on the measures of speech rhythm and tempo. Many characteristics could be reliably identified by looking at deviant values of the rhythm metrics. Thus, tokens which are located away from the norm are good indicators of her disordered and non-canonical speech patterns. And both phonetic and phonological factors contribute to her non-canonical rhythmic pattern. Speech characteristics of the speaker include slow, enunciated, and prolonged realization of segments, frequent insertion of pauses, and release of word-final stop consonants, occasional modification of syllable structure via a schwa-like vowel insertion, fully aspirated stop in the intervocalic context, and a variant realization of [aj] to [a]. Some of these non-canonical phonetic properties result in decreased rate of speech, and result in increased the variability of the consonantal, rather than vocalic, interval, contributing to the characteristics of stress-timing. These characteristics of disordered speech pose a challenge on models of speech production that does not take into account possible modifications of phonetic properties and phonological structure.

5. Acknowledgements

This work is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. Statements in this paper reflect the opinions and conclusions of the authors, and are not necessarily endorsed by the NSERC.

6. References

[1] Louch, M. "A Case of Foreign Accent Syndrome," B.Sc. Thesis. McMaster University, 2009.

[2] Christoph D.H., de Freitas G.R., Dos Santos D.P., Lima M.A., Araújo A.Q., and Carota A., "Different perceived foreign accents in one patient after prerolandic hematoma", Eur. Neurol. 52(4):198-201, 2004.

[3] Katz W.F., Garst D.M., and Levitt J. "The role of prosody in a case of foreign accent syndrome (FAS)", Clin Linguist Phon. 22(7):537-66, 2008.

[4] Moen, I. "Foreign accent syndrome: A review of contemporary explanations," Asphasiology, 14, 5-15, 2000.

[5] Varley, R., Whiteside, S., Hammill, C., and Cooper, K. "Phases in speech encoding and foreign accent syndrome," Journal of Neurolinguistics, 19, 356-369, 2006.

[6] Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. "Buckeye Corpus of Conversational Speech (2nd rel.)", Columbus, OH: Department of Psychology, Ohio State University (Distributor), 2007.

[7] Yoon, T. "Capturing Inter-speaker Invariance Using Statistical Measures of Rhythm", Proc. Speech Prosody 2010, 2010

[8] Ramus, F., Nespor, M. and Mehler, J. "Correlates of linguistic rhythm in the speech signal", Cognition 73: 265-292, 1999.

[9] Boersma, P. and Weenink D. Praat: doing Phonetics by Computer (version 5.2), 2010. [downloadable from http://praat.org].

[10] Grabe, E. and E. L. Low . "Durational variability in speech and the rhythm class hypothesis." In C. Gussenhoven and N. Warner (eds.), Papers in Laboratory Phonology 7. Berlin: Mouton de Gruyter. 2002.

[11] White, L. and Mattys, S.L.. "Calibrating rhythm: First language and second language studies." Journal of Phonetics 35, 501-522, 2007.

[12] Dellwo, V., and Wagner, P. "Relations between language rhythm and speech rate.' In Proceedings of the 15th ICPhS, 471-474, 2003.