Gurugubelli, Krishna; Vuppala, Anil; Nonavinakere Prabhakera, Narendra; Alku, Paavo

Duration of the rhotic approximant /ô/ in spastic dysarthria of different severity levels

Published in:
Speech Communication

DOI:
10.1016/j.specom.2020.09.006

Published: 01/12/2020

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Gurugubelli, K., Vuppala, A., Nonavinakere Prabhakera, N., & Alku, P. (2020). Duration of the rhotic approximant /ô/ in spastic dysarthria of dierent severity levels. *Speech Communication*, *125*, 61-68.
https://doi.org/10.1016/j.specom.2020.09.006

Duration of the rhotic approximant /ɹ/ in spastic dysarthria of different severity levels

Krishna Gurugubelli[a,*], Anil Kumar Vuppala[a], N P Narendra[b], Paavo Alku[b]

[a]Speech Processing Laboratory, LTRC, KCIS, International Institute of Information Technology, Hyderabad, India, 500032.
[b]Aalto University, Department of Signal Processing and Acoustics, Espoo, Finland.

Abstract

Dysarthria is a motor speech disorder leading to imprecise articulation of speech. Acoustic analysis capable of detecting and assessing articulation errors is useful in dysarthria diagnosis and therapy. Since speakers with dysarthria experience difficulty in producing rhotics due to complex articulatory gestures of these sounds, the hypothesis of the present study is that duration of the rhotic approximant /ɹ/ distinguishes dysarthric speech of different severity levels. Duration measurements were conducted using the third formant (F3) trajectories estimated from quasi-closed-phase (QCP) spectrograms. Results indicate that the severity level of spastic dysarthria has a significant effect on duration of /ɹ/. In addition, the phonetic context has a significant effect on duration of /ɹ/, the ɪ-r-ɛ context showing the largest difference in /ɹ/ duration between dysarthric speech of the highest severity levels and healthy speech. The results of this preliminary study can be used in the future to develop signal processing and machine learning methods to automatically predict the severity level of spastic dysarthria from speech signals.

Keywords: Rhotic approximant, Dysarthria, Quasi-closed-phase analysis

*Corresponding author
Email addresses: krishna.gurugubelli@research.iiit.ac.in (Krishna Gurugubelli), anil.vuppala@iiit.ac.in (Anil Kumar Vuppala), narendra.prabhakera@aalto.fi (N P Narendra), paavo.alku@aalto.fi (Paavo Alku)

1. Introduction

Dysarthria is a disorder resulting from weaknesses of neuromuscular execution in motor speech production due to brain tumors, brain injury, stroke, cerebral palsy and facial paralysis (Duffy, 2013). It causes defects in the articulation of speech sounds, which reduces the intelligibility of speech (Doyle et al., 1997). Previous acoustic studies have indicated that dysarthric speech shows increased word and syllable durations, slower transitions between phonemes, and reduced overall speech rate, which are all correlated with a reduced range of articulatory movements in patients with spastic dysarthria (Kent et al., 1992). Slower speaking rate and prolongation of syllable duration have been reported in a few previous studies as distinguished features of dysarthric speech. A study of syllable duration showed that speakers with spastic and ataxic dysarthria exhibit greater mean syllable duration compared to healthy controls (Turner and Weismer, 1993). Another study on syllabic timing revealed that the reduced range of articulatory movements gives rise to a prolongation of syllables in speakers with cerebellar dysarthria (Ackermann and Hertrich, 1994). However, it has also been reported that prolongation affects some syllables more than others (Kent et al., 1999).

The articulatory complexity in manner of articulation (MOA) and place of articulation (POA) can affect duration of speech units in dysarthria (Van Nuffelen et al., 2009). A study on tongue tip kinematic deviations in speakers with spastic dysarthria showed a reduced range of movements during the production of alveolar sounds (Kim et al., 2010b). Moreover, the study on American English native speakers revealed that POAs such as alveolar, post-alveolar, and palatal-alveolar are difficult articulatory gestures for speakers with cerebral palsy (Kim et al., 2010a). The investigation by Kim et al. (2010a) prompted the present authors to study the effects of dysarthria *severity* in the production of alveolar sounds, more specifically, the rhotic approximant /ɹ/. The hypothesis of this study is that the severity level of spastic dysarthria is reflected by the duration of the rhotic approximant /ɹ/ due to complex articulatory gestures required in production of rhotic sounds. It is expected that the duration of the rhotic approximant /ɹ/ increases as a function of dysarthria severity. In addition, the relative change of duration between dysarthric speech of different severity levels and healthy speech is expected to be more prominent in the rhotic approximant /ɹ/ compared to the word overall duration. This study of duration of the rhotic approximant /ɹ/ is a preliminary step in efforts to get better acoustic measures for automatic prediction of the dysarthria severity level from speech signals.

In normal speech, segmental duration depends on several factors such as phonetic context, lexical stress, phrase boundaries, speaking rate and gender (van Santen and Olive, 1990; Tsao and Weismer, 1997; Robb et al., 2005; Van Borsel and De Maesschalck, 2008). The investigations by Simpson (2001, 2009) indicated that gender could be an important factor in accounting for changes in duration of speech sounds and that due to the differences in vocal tract cross-section, articulatory distances required to attain phonetics targets are different between males and females. On the other hand, previous studies investigating the effect of phonetic context on duration of speech segments have revealed that phonetic context alters duration of sound units (McCauley and Skenes, 1987; Jongman, 1989; Mendoza et al., 2003; Koenig, 2007). In particular, the studies by Lockenvitz et al. (2015) and Narayanan et al. (1999) showed that segmental duration of rhotic sounds is longer in vocalic contexts than in consonantal contexts. From the above studies, it can be concluded that gender and phonetic context have been shown to affect duration of the rhotic approximant in normal speech. Therefore, the present study considers gender and phonetic context along with dysarthria severity as factors to investigate duration of the rhotic approximant /ɹ/.

Several studies have shown that word durations are longer in utterances spoken by speakers with spastic dysarthria compared to healthy talkers (Kent et al., 1979; Turner and Weismer, 1993; Ackermann and Hertrich, 1994). Moreover, the study by Liss et al. (2009) indicated that duration of vocalic and consonantal segments is longer in speakers with spastic dysarthria than in speakers with Parkinsonian or ataxic dysarthria. The study by Lee and Hustad (2013) revealed that duration of speech units in children with cerebral palsy increases with the severity of dysarthria. Previous investigations on spastic dysarthria have focused on the relationship between dysarthria severity and duration of the overall utterance (e.g. word duration). In contrast to this, the present study addresses the effect of dysarthria severity on duration of the alveolar sound /ɹ/. Duration of /ɹ/ is studied using the third formant (F3) trajectory estimated from speech. The F3 trajectory is estimated from the spectrogram computed using quasi-closed-phase (QCP) analysis that has been shown to be an accurate method to estimate formants (Airaksinen et al., 2014). Effects of three factors (speaker gender, dysarthria severity, phonetic context of /ɹ/) are analyzed on duration of /ɹ/ using dysarthric speech data of the UA-Speech database (Kim et al., 2008). In addition, prolongation of /ɹ/ duration due to dysarthria is compared to prolongation of word duration.

2. Studies on /ɹ/ in healthy speakers

Alveolar approximants produced by American English healthy speakers have been studied widely in the literature (Espy-Wilson et al., 2000; Zhou et al., 2007; Arai, 2014). Based on the POA and MOA, the rhotic sounds are broadly classified into trills, approximants, taps, flaps, and fricatives. In American English dialects, the most prevalent rhotic sound is the rhotic approximant /ɹ/ (Alwan et al., 1997). Different articulatory configurations involved in the production of rhotic sounds in American English have been investigated using magnetic resonance imaging (MRI), electropalatography (EPG) and ultrasound (Zhou et al., 2007; Alwan et al., 1997; Tiede et al., 2004). These studies have classified the articulatory gestures of rhotics broadly into alveolar, retroflex and bunched articulations. The wide range of articulatory configurations of rhotic sounds have been shown to exhibit articulatory-to-acoustic correlations in the third formant (F3) dynamics (Boyce and Espy-Wilson, 1997; Espy-Wilson et al., 2000; Arai, 2014). Furthermore, durations of American English rhotic sounds have been studied using F3 trajectories obtained from speech signals (Boyce and Espy-

Wilson, 1997; Espy-Wilson et al., 2000). Studies on rhotics are not limited to articulatory-acoustic analysis. Harper et al. (2016) studied L2 speakers' production and acquisition of the American English rhotic approximant. Similarly, the effect of dialects in the acquisition and production of rhotic sounds was investigated by Leemann et al. (2018). Another study in rhotics revealed that the /ɹ/ sounds are difficult to acquire for Japanese children, and that these sounds show wide pronunciation variations (in POA and MOA) (Arai, 2013). Moreover, based on variations in the production of rhotics, bilingual speakers' dominant language has been examined (Kaland et al., 2016). In contrast to the above studies, the present study investigates how duration of the most prevalent rhotic in American English, the rhotic approximant /ɹ/, is affected in spastic dysarthria by the disease severity.

3. Articulation and formants of /ɹ/

3.1. Articulation

A study on American English /ɹ/ reported six different articulatory configurations in the production of rhotic sounds (Delattre and Freeman, 1968). These are broadly classified into alveolar, post-alveolar, and bunched articulations (Alwan et al., 1997; Harper et al., 2016; Zhou et al., 2007; Arai, 2014). In both alveolar and post-alveolar articulations, two constrictions are formed. The first constriction is formed in the lower pharynx region using tongue root and the second constriction is formed in respective places of articulation with either tongue-tip or tongue-blade (Zhou et al., 2007). The /ɹ/ sound produced in bunched articulation has constriction in the lower pharynx and center of the palate with no tongue tip/blade raising (Ladefoged and Maddieson, 1996; Harper et al., 2016). The /ɹ/ sound produced in retroflex articulation shows constriction in lower pharynx and lip rounding. The complex interaction between pharyngeal constriction, labial constriction, and different types of tongue constrictions during the production of /ɹ/ makes it difficult to predict the articulatory gesture from the acoustic speech signal (Ladefoged and Maddieson, 1996; Alwan et al., 1997).

3.2. Formants

The most prominent feature of rhotic sounds in American English is their low F3, which can vary between 1300 Hz and 2100 Hz (Espy-Wilson et al., 2000). In general, F3 of rhotic sounds is around 2000 Hz for both men and women. In other phonemes of American English, F3 is typically located between 2200 Hz and 3000 Hz (Espy-Wilson et al., 2000; Boyce and Espy-Wilson, 1997). The lowering of F3 in /ɹ/ is due to the constriction in pharynx and palate-alveolar region. The other characteristics in the formant structure of the rhotic approximant are lowering of the first formant (F1) and the second formant (F2), and the proximity between F2 and F3 (Alwan et al., 1997; Delattre and Freeman, 1968). During speech production, the phonemic boundaries of the /ɹ/ sounds are manifested in F3 transitions around 2000 Hz (Ladefoged and Maddieson, 1996; Boyce and Espy-Wilson, 1997). In a vowel-rhotic-vowel context, the F3 trajectory attains a parabolic shape. Similarly in a silence-rhotic-vowel context (where the rhotic sound is the starting utterance), raising of F3 is manifested (Boyce and Espy-Wilson, 1997; Espy-Wilson et al., 2000).

Among the different acoustic correlates of the rhotic approximant, the lowering of F3 is preserved in different articulatory configurations (alveolar, retroflex, and bunched) during the production of /ɹ/ (Ladefoged and Maddieson, 1996). Rhotic sounds can be detected by analysing dips in F3 trajectories (Espy-Wilson, 1994). Based on above studies, the present study defines duration of /ɹ/ as the time-span when the F3 trajectory is below 2000 Hz.

4. The experimental setup

In order to estimate the duration of /ɹ/ using the F3 trajectory, this study takes advantage of spectrograms computed with QCP analysis (Airaksinen et al., 2014). Duration of /ɹ/ is measured from healthy and dysarthric speech of the UA-Speech database (Kim et al., 2008). The UA-Speech database provides information about three attributes (speaker gender, dysarthria severity, phonetic context of /ɹ/) whose impact on duration of /ɹ/ is investigated. In the following sub-sections, QCP analysis, the UA-Speech database and the procedure to estimate duration of /ɹ/ are described in more detail.

4.1. QCP analysis

QCP analysis is a recently proposed speech analysis method that has been developed particularly for glottal inverse filtering and formant estimation (Airaksinen et al., 2014; Gowda et al., 2016). QCP computes all-pole models of the vocal tract using weighted linear prediction analysis with the attenuated main excitation (AME) weighting function (Alku et al., 2013). The AME weighting function enables reducing the contribution of the voice source in the computation of the speech signal's autocorrelation function. Hence, QCP analysis attenuates the coupling between the vocal tract and glottal source, leading to formant estimates that are more accurate than those computed by, for example, conventional linear prediction (Airaksinen et al., 2014; Gowda et al., 2016; Alku et al., 2013).

Figure 1 demonstrates QCP analysis by depicting an utterance with /ɹ/ (word *zero*) both as a time-domain waveform (the upper pane) and as a QCP spectrogram (the lower pane). In the QCP spectrogram, the arrow with diamond head indicates the F3 trajectory. In addition, the two vertical dashed arrows show the boundaries of /ɹ/ marked by the experimenter. In defining the arrow positions, the experimenter inspected visually the F3 trajectory from the QCP spectrogram by searching for a time-span over which F3 was below the selected limit of 2000 Hz.

4.2. The UA-Speech database

The UA-Speech corpus is a widely used publically available database that includes speech recorded from 16 speakers (four females) with dysarthria, diagnosed with cerebral palsy and spastic dysarthria and 13 healthy controls (four females). From each speaker, 765 isolated word utterances (300 distinct uncommon words and three replications of radio alphabet, digits, computer commands,

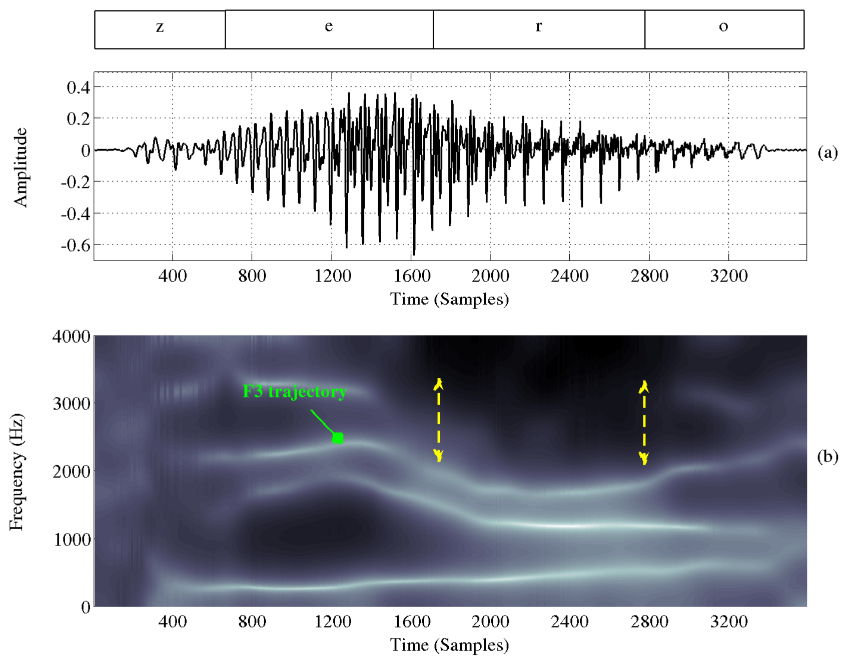


Figure 1: (Color online) QCP analysis for an utterance (word *zero*) from the UA-Speech database: (a) speech signal and (b) QCP spectrogram.

and common words) were collected with a microphone array. As a part of the database design, speech intelligibility was measured by asking native listeners to transcribe words produced by the speakers with dysarthria. The percentage of correct responses, averaged over five native listeners, were calculated. Based on the average correct responses (in %), the speakers with dysarthria of the UA-Speech database are divided into four severity categories according to Table 1.

Table 1: Classification of the speakers with dysarthria in the UA-Speech database

Average % of correct responses in the subjective evaluation	Intelligibility level	Severity level	Number of speakers
0-25 %	very low	very high	4(3 males, 1 female)
26-50 %	low	high	4 (3 males, 1 female)
51-75 %	medium	medium	3 (2 males, 1 female)
76-100 %	high	low	5 (4 males, 1 female)
100 %	very high	healthy	13 (9 males, 4 females)

The number of speakers in each severity level is tabulated in Table 1. Among the four dysarthric female speakers of UA-Speech, there is one speaker for each severity level. From the 12 dysarthric male speakers, there are four, two, three and three speakers, respectively, for severity levels low, medium, high and very high. Out of all utterances, the present study focuses only on those words that include the rhotic approximant /ɹ/. The chosen words, their phonetic description and phonetic context of /ɹ/ are tabulated in Table 2.

4.3. Estimation of duration of /ɹ/

Duration of /ɹ/ was measured by visually observing the QCP spectrogram for each word that included the rhotic approximant /ɹ/. As mentioned in Section 3.2, the F3 trajectory of /ɹ/ shows typically a shape of a parabola in a vowel-rhotic-vowel context as demonstrated schematically in Figure 2. For each word, duration of /ɹ/ was defined from the QCP spectrogram as the time-span (shown by the arrow in Figure 2) when F3 trajectory is below 2000 Hz.

Table 2: List of words and phonetic contexts of /ɹ/ used in the present study.

Word	Phonetic description	Phonetic context of /ɹ/
zero	zɪro	ɪ-r-o
paragraph	ˈpɛrəˌgræf	ɛ-r-ə and ɡ-r-æ
sierra	siɛrə	ɛ-r-ə
irresolute	ɪˈrɛzəlːut	ɪ-r-ɛ
burrows	bəroz	ə-r-o
right	rajt	r-a
roof	ruf	r-u

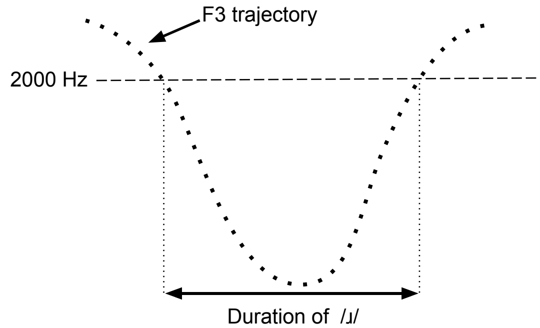


Figure 2: An illustration of using the F3 trajectory in measuring duration of /ɹ/.

In measuring /ɹ/ duration values from QCP spectrograms, F3 was easily observed for most of the utterances. However, F3 lowering was not observed in a few tokens in high and very high severity levels of dysarthria. Furthermore, F3 lowering was observed between 2000 Hz and 2200 Hz in a few tokens. In total, 28 tokens (22 and 6 tokens by male and female speakers, respectively) were excluded from further analyses. The number of tokens analysed in each dysarthria severity category and in each phonetic context is shown in Table 3 and Table 4, respectively, together with the number of tokens removed in each category.

The duration measurement procedure described above might be affected by the subjective criteria of the experimenter. Therefore, the procedure was tested in terms of its interjudge and intrajudge reliability. For the former, five ex-

Table 3: The number of tokens analysed (#A) and excluded (#E) from the analyses in each dysarthria severity category.

Severity category	**#A**	**#E**
healthy	233	1
low	86	4
medium	54	0
high	67	5
very high	54	18
Total	**494**	**28**

Table 4: The number of tokens analysed (#A) and excluded (#E) from the analyses in each phonetic context.

Phonetic context	**#A**	**#E**
PC1: ɛ-r-ə (paragraph)	81	6
PC2: ɛ-r-ə (sierra)	81	6
PC3: ɪ-r-o (zero)	82	5
PC4: ɡ-r-æ (paragraph)	81	6
PC5: ɪ-r-ɛ (irresolute)	28	1
PC6: r-a (right)	85	2
PC7: r-u (roof)	28	1
PC8: ə-r-o (burrows)	28	1
Total	**494**	**28**

perimenters measured /ɹ/ durations using the procedure described. For the latter, one of the experimenters repeated the duration analysis five times over twenty days. To assess the measurement reliability, statistical analyses were computed with one-way analysis of variances (ANOVA) by using duration of /ɹ/ as dependent variable and by using experimenter and measurement as independent variable in testing interjudge and intrajudge reliability, respectively. Further, Pearson correlation coefficient (ρ) was computed for the data of both the interjudge and intrajudge test. The ANOVA conducted on the interjudge data showed no statistically significant differences between the experimenters [df = 4, F = 0.0614, and p = 0.804]. Similarly, the ANOVA computed from the intrajudge data indicated no statistical differences between the measurements [df = 4, F = 0.0018, and p = 0.9664]. Moreover, strong correlation values were obtained both for the interjudge ($\rho = 0.939$) and intrajudge ($\rho = 0.977$) measurements. Taken together, these statistical tests indicated that the proposed subjective measurement procedure is reliable to be used to measure duration of /ɹ/.

5. Results

Previous studies have shown that word durations are longer in dysarthric speech compared to healthy speech (Turner and Weismer, 1993; Ackermann and Hertrich, 1994). Moreover, word duration is more straightforward to be measured than /ɹ/ duration. Therefore, an experiment was conducted in the current study to first study whether duration of /ɹ/ is capable of better indicating changes in the severity level of dysarthria compared to word duration. If duration of /ɹ/ turned out to be a better indicator of the dysarthria severity, further statistical analyses were computed to study the effects of three factors (speaker gender, dysarthria severity, phonetic context of /ɹ/) on this duration measure. The statistical analyses were conducted using the R-tool kit (Team et al., R Foundation for Statistical Computing, Vienna, Austria, 2013).

To compare the duration measures, two one-way ANOVAs were first computed by using dysarthria severity (healthy, low, medium, high, very high) as independent variable and the corresponding duration measure as dependent variable. The ANOVA analyses indicated that dysarthria severity had a significant effect on both word duration [df = 4, F = 8.2095, p = 2.014e-05] and on /ɹ/ duration [df = 4, F = 77.73, $p < 2.2e\text{-}16$]. Furthermore, the average duration in the rhotic approximant and words is shown in Figure 3 as function of the dysarthria severity level. This figure demonstrates that the average duration increases both in the rhotic and in words as a function of the dysarthria severity level (except in words between healthy speakers and speakers with dysarthria of the lowest severity level). However, it was found that the relative change in /ɹ/ duration between dysarthric speech and healthy speech was larger than the corresponding relative change in word duration. The relative change in /ɹ/ duration ($D_{/ɹ/}$) is computed as

$$D_{/ɹ/} = \frac{RD_s}{RD_h}, \tag{1}$$

where RD_s denotes the average /ɹ/ duration of dysarthric speech of a specific severity level (low/medium/high/very high) and RD_h denotes the average /ɹ/ duration of healthy speech. Similarly, the relative change in word duration (D_w) is computed as

$$D_w = \frac{WD_s}{WD_h}, \tag{2}$$

where WD_s denotes the average word duration in dysarthric speech of a specific severity level (low/medium/high/very high) and WD_h denotes the average word duration in healthy speech. $D_{/ɹ/}$ and D_w are shown in Figure 4 with gray and black bars, respectively, in the four severity levels of dysarthria. This figure indicates that duration of /ɹ/ reflects changes in dysarthria severity more clearly than word duration.

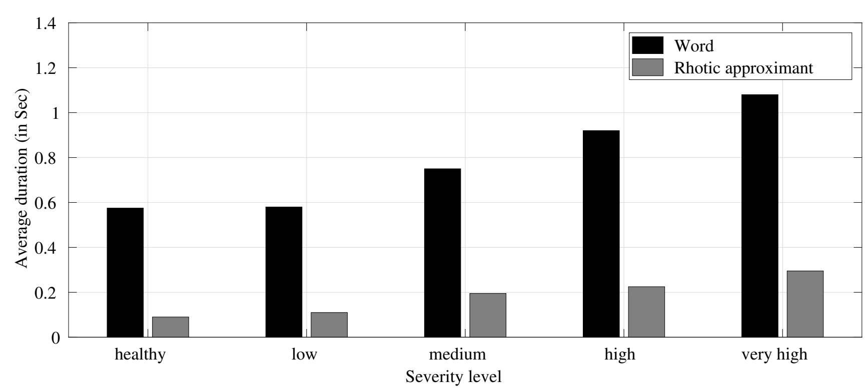


Figure 3: Average duration of words (black bars) and /ɹ/ (gray bars) as a function of dysarthria severity level.

Furthermore, this study investigated the effect of two between-subject factors (gender and severity) and one within-subject factor (phonetic context) on rhotic

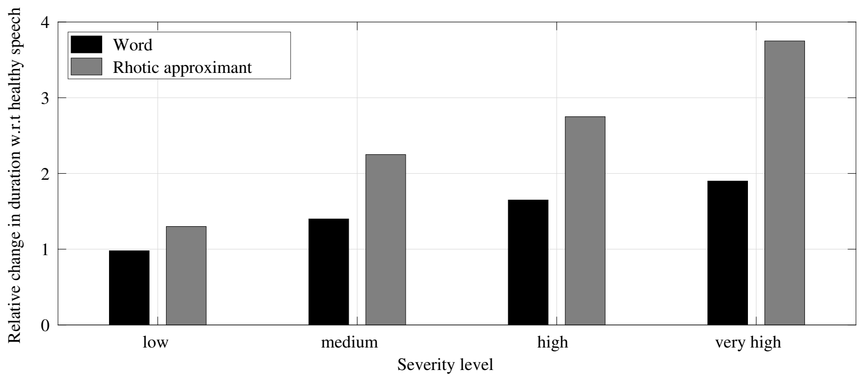


Figure 4: Relative change of duration between dysarthric and healthy speech for words (black bars), defined by D_w in Eq. 2, and for /ɹ/ (gray bars), defined by $D_{/ɹ/}$ in Eq. 1, as a function of dysarthria severity level.

approximant duration. In order to study how duration of /ɹ/ is affected in dysarthria by the three factors, a three-way ANOVA was computed as in many similar studies (Jacewicz et al., 2010, 2007; Boyce and Espy-Wilson, 1997). The ANOVA analysis was computed using duration of /ɹ/ as dependent variable and the gender (male, female), severity (healthy, low, medium, high, very high), and phonetic context (ɪ-r-o, ɛ-r-ə (paragraph), ɡ-r-æ, ɛ-r-ə (sierra), ɪ-r-ɛ, ə-r-o, r-a, and r-u) as independent variables. Pairwise post-hoc analyses were conducted with Tukey's honestly significant difference tests. The 3-way ANOVA analysis showed that duration of /ɹ/ was influenced by the gender [df = 1, F = 16.76, p=0.0003], the severity [df = 4, F = 262.07, p << 0.0001], and the phonetic context [df = 7, F = 13.47, p << 0.0001]. In addition, the interaction between the severity and gender [df = 4, F = 7.99, p=0.0002] and between the severity and phonetic context [df = 28, F = 2.63, p=0.006] affected duration of /ɹ/ significantly. The interaction between the gender and the phonetic context [df = 7, F = 1.189, p=0.34] was, however, not significant. Further, the post-hoc tests of the severity indicated that all pairwise comparisons between the severity classes were significantly different (p < 0.0001) except between healthy and low (p = 0.0719), and between medium and high (p = 0.001). Finally, the mean duration of /ɹ/ was significantly lower for female speakers compared to male speakers.

Post-hocs of the interactions between the severity and phonetic context showed that the number of significantly different ($p < 0.0001$) pairwise comparisons was largest for the pairs where either ɛ-r-ə or ɪ-r-ɛ was compared to another class of phonetic context. Based on the neighbouring phones of the rhotic approximant, the phonetic contexts were grouped into three categories: vowel-rhotic-vowel (VRV), consonant-rhotic-vowel (CRV) and rhotic-vowel (RV). The effect of the neighbouring phone on /ɹ/ duration was studied and it was observed that the rhotic sound produced in the VRV context showed a higher mean duration (194 ms) compared to the RV (157 ms) and CRV (152 ms) context. Furthermore, the effect of the neighbouring phone in different severity levels of dysarthria was studied. The mean duration of the rhotic sound as a function of the severity level is tabulated in Table 5 for all the three neighboring phone contexts. The table shows that the mean duration of /ɹ/ is higher in the VRV context compared to the RV and CRV contexts in all severity levels. This result is graphically depicted in Figure 5 with the help of $D_{/ɹ/}$ defined in Eq. 1.

Table 5: Duration of the rhotic approximant (in ms) as a function of dysarthria severity level for the three neighbouring phone contexts.

Severity level	Neighbouring phone context		
	VRV	RV	CRV
healthy	91	66	60
low	112	84	93
medium	217	135	131
high	238	188	195
very high	310	299	290

The measured duration values are shown in Figure 6 as a function of the analysed factor. In Figure 6(a), duration is depicted for female and male speakers separately in the five severity levels. Figure 6(b) shows duration of /ɹ/ as a

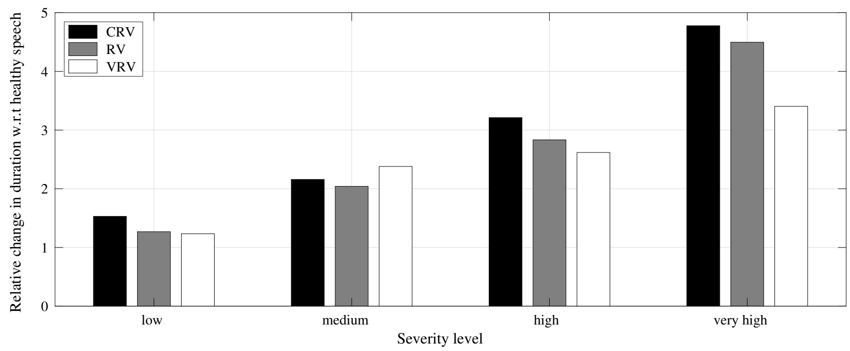


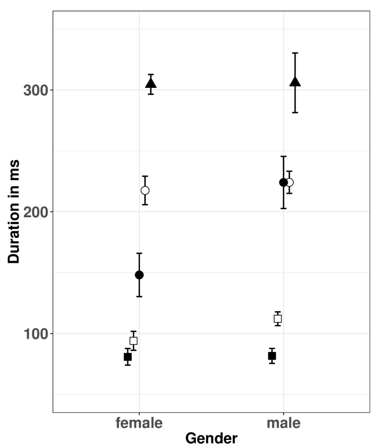
Figure 5: Relative change in /ɹ/ duration (defined in Eq. 1) in the three different neighbouring phone contexts as a function of severity level.

function of the severity level separately for each phonetic context. Furthermore, duration of /ɹ/ is shown as a function of the phonetic context separately for female and male speakers in Figure 6(c).

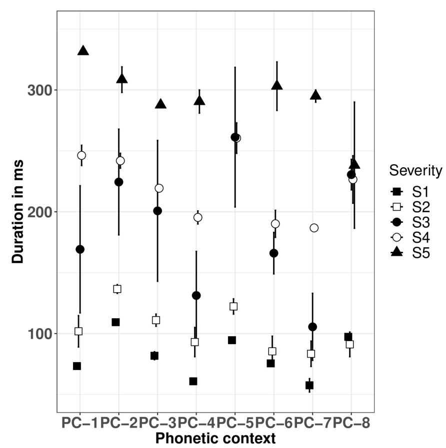
Finally, examples of QCP spectrograms computed from speech utterances (word *sierra*) taken from the UA-Speech database are shown in Figure 7. The figure shows spectrograms computed from a healthy speaker and from speakers with dysarthria in each severity level. For each speech utterance, the figure shows the measured duration of the rhotic approximant /ɹ/. The figure demonstrates, for example, that the duration of the rhotic approximant /ɹ/ increases from a low value (about 95 ms) to a high value (about 335 ms) when healthy speech is compared to dysarthric speech of a very high severity level.

6. Discussion

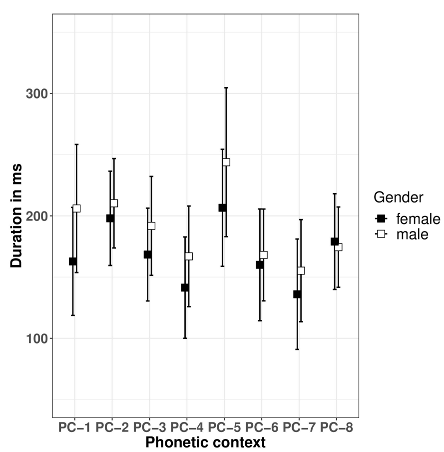
The present study investigates duration of the rhotic approximant /ɹ/ in dysarthric speech in American English as a function of dysarthria severity, phonetic context, and gender. The study showed that duration of /ɹ/ was significantly higher in dysarthric speech compared to healthy speech. In addition, it was shown that the relative increase in duration of /ɹ/ due to dysarthria was larger than the corresponding relative duration increase in words. Duration of



(a)



(b)



(c)

Figure 6: Duration of /ɹ/ as a function of gender (a), severity level (b) and (c) phonetic context. The dots (square, circular, triangular, rhombus) and whiskers denote mean values and standard deviations, respectively. Severity levels are denoted by S1 (healthy), S2 (low), S3 (medium), S4 (high), and S5 (very high). Phonetic contexts are denoted by PC-1 (ɛ-r-ə) in *paragraph*, PC-2 (ɛ-r-ə) in *sierra*, PC-3 (ɪ-r-o), PC-4 (g-r-æ), PC-5 (ɪ-r-ɛ), PC-6 (r-a), PC-7 (r-u), and PC-8 (ə-r-o).

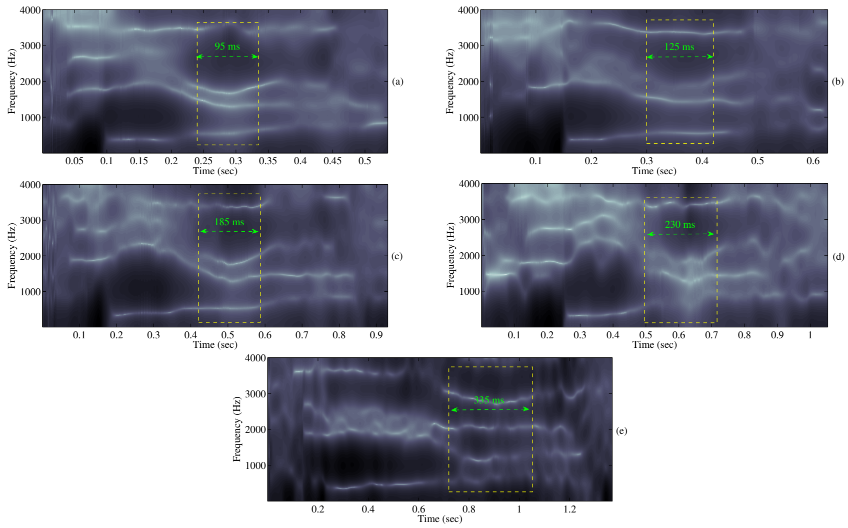


Figure 7: QCP analysis for an utterance (word *sierra*) from the UA-Speech database. QCP spectrograms are shown for utterances spoken by a healthy speaker (a) and by speakers with spastic dysarthria of four severity levels: (b) low, (c) medium, (d) high, and (e) very high. The estimated /ɹ/ durations are shown by green arrows.

the rhotic approximant produced in intervocalic (vowel-rhotic-vowel) phonetic contexts was higher than that in other phonetic contexts. These results are described in greater detail below.

As demonstrated in Figure 6(a), the mean duration of /ɹ/ was lowest for healthy speakers and it increased both for female and male subjects when the severity level of dysarthria changed from low to very high. This finding is in line with previous studies which have reported increased syllable durations in dysarthria (Turner and Weismer, 1993) as well as slower transitions between phonemes and reduced speech rate in patients with spastic dysarthria (Kent et al., 1992). The significant effect of gender on duration is explained mainly by the differences in the low and medium levels of dysarthria for which males showed larger duration values than females, while in the other severity classes the differences in duration were small between genders. The significant effect of the severity level on duration of /ɹ/ is demonstrated by Figure 6(b) which shows clearly an increasing trend in duration as a function of the severity level: while the mean durations in healthy speech are approximately between 50 ms and 110 ms, they vary between 240 ms and 380 ms in very high severity. It is though worth observing that the increase in duration was small between healthy speakers and speakers with low severity level of dysarthria, but the duration values rose extensively for speakers with more severe levels of the disease. The joint effect of the severity level and phonetic context on duration of /ɹ/ can be demonstrated by comparing Figure 6(b) and Figure 6(c).

From the results, it can be seen that two phonetic contexts (ɛ-r-ə and particularly ɪ-r-ɛ) show large values of duration. As it can be seen in Figure 6(a), these two phonetic categories were associated with the longest mean /ɹ/ durations in the two most severe classes of dysarthria. These phonetic categories belong to the VRV context. From Table 5, it can be observed that the mean duration of rhotic approximant sounds produced in the VRV context is higher than that in the RV and CRV context. This result is line with the studies of Lockenvitz et al. (2015) and Narayanan et al. (1999) which reported that duration of rhotic sounds produced in the vocalic context was higher than that in the consonantal phonetic context. From Figure 5, it can be seen that the relative change in duration of rhotic sounds is larger in the CRV context than in the VRV context. This is in line with results of Kent et al. (1979), which showed that the slope of formant transition is low in consonant-vowel regions due to slower articulatory dynamics.

In summary, the results suggest that duration of /ɹ/, measured using F3 trajectories extracted from QCP spectrograms, is capable of distinguishing dysarthric speech of different severity levels: the longer the duration, the more severe is the level of dysarthria in general. An exception to this general trend is the lowest severity level (also referred as the early stage of dysarthria) which shows overlapped duration values with those measured from healthy speakers. Interestingly, the phonetic context of /ɹ/ showed a significant effect on duration of /ɹ/ revealing that particularly the ɪ-r-ɛ context was effective in distinguishing dysarthric speech from utterances spoken by healthy talkers.

7. Conclusion

Duration of the American English rhotic approximant /ɹ/ was studied in this investigation in the context of spastic dysarthria severity. Duration values were estimated from F3 trajectories that were obtained by analysing visually QCP spectrograms using a known, straightforward criterion in rhotic detection. The analysis was conducted for dysarthric and healthy speech using the UA-Speech database. The study focused on the effects of three factors (gender, dysarthria severity, phonetic context) on duration of /ɹ/. The results showed that duration of /ɹ/ is an effective indicator of the severity level of dysarthria except between utterances of the lowest level of dysarthria and utterances of healthy speakers. Moreover, among the analysed phonetic contexts, the ɪ-r-ɛ context showed the largest differences in duration of /ɹ/ between the most severe levels of dysarthria and healthy speech.

The current study has resulted in valuable new information about duration of the rhotic approximant /ɹ/ in dysarthric speech of different severity levels. This preliminary study will be followed by future investigations in which signal processing and machine learning methods will be developed to automatically detect duration of /ɹ/ from QCP spectrograms to be used in the automatic, speech-based severity assessment of dysarthric speech.

References

Ackermann, H., Hertrich, I., 1994. Speech rate and rhythm in cerebellar dysarthria: An acoustic analysis of syllabic timing. Folia Phoniatrica et Logopaedica 46, 70–78.

Airaksinen, M., Raitio, T., Story, B.H., Alku, P., 2014. Quasi closed phase glottal inverse filtering analysis with weighted linear prediction. IEEE/ACM Transactions on Audio, Speech, and Language Processing 22, 596–607.

Alku, P., Pohjalainen, J., Vainio, M., Laukkanen, A.M., Story, B.H., 2013.

Formant frequency estimation of high-pitched vowels using weighted linear prediction. The Journal of the Acoustical Society of America 134, 1295–1313.

Alwan, A., Narayanan, S., Haker, K., 1997. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. The Journal of the Acoustical Society of America 101, 1078–1089.

Arai, T., 2013. On why Japanese /r/ sounds are difficult for children to acquire, in: Proc. Interspeech, pp. 2445–2449.

Arai, T., 2014. Retroflex and bunched English /r/ with physical models of the human vocal tract, in: Proc. Interspeech, pp. 706–710.

Boyce, S., Espy-Wilson, C.Y., 1997. Coarticulatory stability in American English /r/. The Journal of the Acoustical Society of America 101, 3741–3753.

Delattre, P., Freeman, D.C., 1968. A dialect study of American r's by X-ray motion picture. Linguistics 6, 29–68.

Doyle, P.C., Leeper, H.A., Kotler, A.L., Thomas-Stonell, N., O'Neill, C., Dylke, M.C., Rolls, K., 1997. Dysarthric speech: a comparison of computerized speech recognition and listener intelligibility. Journal of Rehabilitation Research and Development 34, 309–316.

Duffy, J.R., 2013. Motor Speech Disorders: Substrates, Differential Diagnosis, and Management. 3rd ed. Elsevier Health Sciences.

Espy-Wilson, C.Y., 1994. A feature-based semivowel recognition system. The Journal of the Acoustical Society of America 96, 65–72.

Espy-Wilson, C.Y., Boyce, S.E., Jackson, M., Narayanan, S., Alwan, A., 2000. Acoustic modeling of American English /r/. The Journal of the Acoustical Society of America 108, 343–356.

Gowda, D., Airaksinen, M., Alku, P., 2016. Quasi closed phase analysis of speech signals using time varying weighted linear prediction for accurate formant tracking, in: Proc. ICASSP, pp. 4980–4984.

Harper, S., Goldstein, L., Narayanan, S.S., 2016. L2 acquisition and production of the English rhotic pharyngeal gesture, in: Proc. Interspeech, pp. 208–212.

Jacewicz, E., Fox, R.A., Salmons, J., 2007. Vowel duration in three american english dialects. American Speech 82, 367–385.

Jacewicz, E., Fox, R.A., Wei, L., 2010. Between-speaker and within-speaker variation in speech tempo of American English. The Journal of the Acoustical Society of America 128, 839–850.

Jongman, A., 1989. Duration of frication noise required for identification of English fricatives. The Journal of the Acoustical Society of America 85, 1718–1725.

Kaland, C., Galatà, V., Spreafico, L., Vietti, A., 2016. /r/ as language marker in bilingual speech production and perception, in: Proc. Interspeech, pp. 515–519.

Kent, J.F., Kent, R.D., Rosenbek, J.C., Weismer, G., Martin, R., Sufit, R., Brooks, B.R., 1992. Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis. Journal of Speech, Language, and Hearing Research 35, 723–733.

Kent, R.D., Netsell, R., Abbs, J.H., 1979. Acoustic characteristics of dysarthria associated with cerebellar disease. Journal of Speech, Language, and Hearing Research 22, 627–648.

Kent, R.D., Weismer, G., Kent, J.F., Vorperian, H.K., Duffy, J.R., 1999. Acoustic studies of dysarthric speech: Methods, progress, and potential. Journal of Communication Disorders 32, 141–186.

Kim, H., Hasegawa-Johnson, M., Perlman, A., Gunderson, J., Huang, T.S., Watkin, K., Frame, S., 2008. Dysarthric speech database for universal access research, in: Proc. Interspeech, pp. 1741–1744.

Kim, H., Martin, K., Hasegawa-Johnson, M., Perlman, A., 2010a. Frequency of consonant articulation errors in dysarthric speech. Clinical Linguistics & Phonetics 24, 759–770.

Kim, H., Rong, P., Loucks, T.M., Hasegawa-Johnson, M., 2010b. Kinematic analysis of tongue movement control in spastic dysarthria, in: Proc. Interspeech, pp. 2578–2581.

Koenig, L., 2007. Characteristics of /h/ in American English: Effects of speaker and phonetic context. The Journal of the Acoustical Society of America 122, 2973–2973.

Ladefoged, P., Maddieson, I., 1996. The Sounds of the World's Languages. 1st ed. Blackwell Oxford.

Lee, J., Hustad, K.C., 2013. A preliminary investigation of longitudinal changes in speech production over 18 months in young children with cerebral palsy. Folia Phoniatrica et Logopaedica 65, 32–39.

Leemann, A., Schmid, S., Studer-Joho, D., Kolly, M.J., 2018. Regional variation of /r/ in Swiss German dialects, in: Proc. Interspeech, pp. 2738–2742.

Liss, J.M., White, L., Mattys, S.L., Lansford, K., Lotto, A.J., Spitzer, S.M., Caviness, J.N., 2009. Quantifying speech rhythm abnormalities in the dysarthrias. Journal of Speech, Language, and Hearing Research 52, 1334–1352.

Lockenvitz, S., Kuecker, K., Ball, M.J., 2015. Evidence for the distinction between "consonantal-/r/" and "vocalic-/r/" in American English. Clinical Linguistics & Phonetics 29, 613–622.

McCauley, R.J., Skenes, L.L., 1987. Contrastive stress, phonetic context, and misarticulation of /r/ in young speakers. Journal of Speech, Language, and Hearing Research 30, 114–121.

Mendoza, E., Carballo, G., Cruz, A., Fresneda, M.D., Muñoz, J., Marrero, V., 2003. Temporal variability in speech segments of spanish: Context and speaker related differences. Speech Communication 40, 431–447.

Narayanan, S., Byrd, D., Kaun, A., 1999. Geometry, kinematics, and acoustics of tamil liquid consonants. The Journal of the Acoustical Society of America 106, 1993–2007.

Robb, M., Gilbert, H., Lerman, J., 2005. Influence of gender and environmental setting on voice onset time. Folia Phoniatrica et Logopaedica 57, 125–133.

van Santen, J.P., Olive, J.P., 1990. The analysis of contextual effects on segmental duration. Computer Speech & Language 4, 359–390.

Simpson, A.P., 2001. Dynamic consequences of differences in male and female vocal tract dimensions. The Journal of the Acoustical Society of America 109, 2153–2164.

Simpson, A.P., 2009. Phonetic differences between male and female speech. Language and Linguistics Compass 3, 621–640.

Team, R.C., et al., R Foundation for Statistical Computing, Vienna, Austria, 2013. R: a language and environment for statistical computing .

Tiede, M.K., Boyce, S.E., Holland, C.K., Choe, K.A., 2004. A new taxonomy of American English /r/ using MRI and ultrasound. The Journal of the Acoustical Society of America 115, 2633–2634.

Tsao, Y.C., Weismer, G., 1997. Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component. Journal of Speech, Language, and Hearing Research 40, 858–866.

Turner, G.S., Weismer, G., 1993. Characteristics of speaking rate in the dysarthria associated with amyotrophic lateral sclerosis. Journal of Speech, Language, and Hearing Research 36, 1134–1144.

Van Borsel, J., De Maesschalck, D., 2008. Speech rate in males, females, and male-to-female transsexuals. Clinical Linguistics & Phonetics 22, 679–685.

Van Nuffelen, G., Middag, C., De Bodt, M., Martens, J.P., 2009. Speech technology-based assessment of phoneme intelligibility in dysarthria. International Journal of Language & Communication Disorders 44, 716–730.

Zhou, X., Espy-Wilson, C.Y., Tiede, M., Boyce, S., 2007. An articulatory and acoustic study of retroflex and bunched American English rhotic sound based on MRI, in: Proc. Interspeech, pp. 54–57.