JSLHR

Research Article

Effects of Phonetic Context on Relative Fundamental Frequency

Yu-An S. Lien,[a] Caitlin I. Gattuccio,[a] and Cara E. Stepp[a]

Purpose: The effect of phonetic context on relative fundamental frequency (RFF) was examined, in order to develop stimuli sets with minimal within-speaker variability that can be implemented in future clinical protocols.
Method: Sixteen speakers with healthy voices produced RFF stimuli. Uniform utterances consisted of 3 repetitions of the same voiced sonorant-voiceless consonant-voiced sonorant speech sequence; moderately variable sentences contained speech sequences with a single voiceless phoneme (/f/, /s/, /ʃ/, /p/, /t/, or /k/); highly variable sentences were loaded with speech sequences using multiple phonemes. Effects of stimulus type (uniform, moderately variable, and highly variable) and phoneme identity (/f/, /s/, /ʃ/, /p/, /t/, and /k/) on RFF means and standard deviations were determined.
Results: Stimulus type and the interaction of vocal cycle and stimulus type were significant for RFF means and standard deviations but with small effect sizes. Phoneme identity and the interaction of vocal cycle and phoneme identity on RFF means and standard deviations were also significant with small to medium effect sizes.
Conclusions: For speakers with healthy voices, uniform utterances with /f/ and /ʃ/ have the lowest standard deviations and thus are recommended for RFF-based assessments. Future work is necessary to extend these findings to disordered voices.

Key Words: voice, speech production, assessment, acoustics

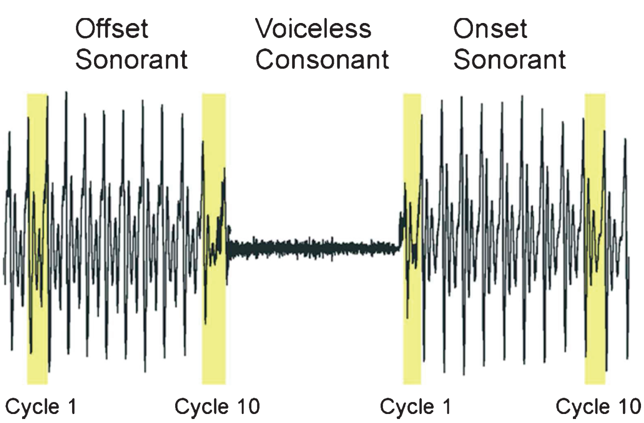
Relative fundamental frequency (RFF) is an acoustic measure that captures the instantaneous changes in fundamental frequency (F0) as a speaker transitions from voicing into and out of a voiceless obstruent (Figure 1). It has been previously measured primarily from voiced sonorant-voiceless consonant-voiced sonorant productions (speech sequence) in running speech (Figure 1) and is defined as the F0s of 10 vocal cycles before and after the voiceless obstruent normalized, in semitones (ST), to relatively steady-state portions of the voicing. Several studies have analyzed the characteristics of RFF (Goberman & Blomgren, 2008; Robb & Smith, 2002; Stepp, 2013; Stepp, Hillman, & Heaton, 2010; Stepp, Merchant, Heaton, & Hillman, 2011; Stepp, Sawin, & Eadie, 2012; Watson, 1998), almost all of which have used a different set of RFF stimuli. Examining the effect of phonetic context on RFF will allow for development of stimuli sets with minimal within-speaker variability that can be implemented in clinical studies and protocols and will further aid in comparisons across studies that use different stimuli sets.

The physiological mechanisms behind RFF have been hypothesized by Watson (1998) and Stepp et al. (2011) to involve the interplay of tension, aerodynamics, and vocal-fold kinematics. Increase in vocal-fold tension, which is achieved in part by the contraction of the cricothyroid muscle, is known to increase F0 (Arnold, 1961; Roubeau, Chevrie-Muller, & Lacau Saint Guily, 1997). During the production of the speech sequences, the activity of the cricothyroid muscle increases preceding or during the voiceless consonant and decreases immediately following the start of the consonant (Löfqvist, Baer, McGarr, & Story, 1989). This increase in tension could potentially lead to an increase in the instantaneous F0s of the vocal cycles surrounding the voiceless consonant. Maximum and minimum airflow during the speech sequence have also been found to be high following voiceless consonants (Löfqvist, Koenig, & Mcgowan, 1995; Löfqvist & Mcgowan, 1992). Higher airflow may cause the vocal folds to be drawn together more quickly, causing the F0s in vocal cycles following the voiceless consonant to increase (Ladefoged, 1972). Finally, vocal-fold abduction has been observed to occur during the offset vowel slightly before the transition into the voiceless consonant (Fukui & Hirose, 1983). Vocal-fold abduction during voicing may increase the durations of both the contacting phase and decontacting phases of the vocal cycles and has been hypothesized to result in lower F0s in the vocal cycles preceding the voiceless

[a]Boston University, Boston, MA
Correspondence to Cara E. Stepp: cstepp@bu.edu
Editor: Jody Kreiman
Associate Editor: Kate Bunton
Received June 20, 2013
Revision received October 1, 2013
Accepted January 6, 2014
DOI: 10.1044/2014_JSLHR-S-13-0158

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

Figure 1. An acoustic waveform of the voiced sonorant-voiceless consonant-voiced sonorant speech sequence /ifi/. The first and 10th vocal cycles for both offset and onset sonorants are highlighted. To calculate relative fundamental frequency (RFF), the instantaneous fundamental frequencies (inverse of the period) of 10 vocal cycles preceding and following the consonant are normalized to the steady-state fundamental frequencies. The instantaneous fundamental frequencies for Offset Vocal Cycle 1 and Onset Vocal Cycle 10 are the steady-state fundamental frequencies for the offset and onset vocal cycles, respectively.



consonant (Watson, 1998). These physiological effects in sum may explain the patterns of RFF noted in healthy young speakers. The effects of tension and vocal-fold kinematics are canceled in the offset, resulting in fairly stable or slightly decreasing offset RFF that reaches a value of 0.44 to –0.84 ST by Offset Vocal Cycle 10 (Robb & Smith, 2002; Watson, 1998). During vowel onset, the effects of tension and aerodynamics sum, causing high RFF values immediately following the consonant starting with initial (Onset Vocal Cycle 1) mean RFF values between 2.3 ST and 3.3 ST (Robb & Smith, 2002; Watson, 1998).

RFF has also been examined in healthy older adults (Watson, 1998). In typical older speakers, offset RFF tends to be lower, reaching a value of –1.66 ST at Vocal Cycle 10; however, onset RFF does not differ significantly between younger and older speakers (Watson, 1998). On the basis of these data, Watson hypothesized that vocal-fold abduction is the primary mechanism for devoicing in older speakers, because morphological and neuromuscular changes in the vocal folds "may limit the ability of aged speakers to produce transient increases in vocal fold tension as part of the devoicing gesture" (Watson, 1998, p. 3646).

Similar to older adults, individuals with voice disorders have also shown patterns of RFF that differ from healthy young adults. For speakers with vocal hyperfunction and Parkinson's disease, both offset and onset RFF tend to be lower compared with age-matched controls (Goberman & Blomgren, 2008; Stepp et al., 2010). Both of these disorders are associated with excessive laryngeal tension, that is, higher baseline tension (Berardelli, Sabra, & Hallett, 1983; Gallena, Smith, Zeffiro, & Ludlow, 2001; Hillman, Holmberg, Perkell, Walsh, & Vaughan, 1989; Roy, Ford, & Bless, 1996), leading to the hypothesis that lower RFF values seen in these individuals may be due to a ceiling effect in which these individuals are not able to create large phoneme-mediated changes in laryngeal tension because of high baseline tension (Stepp et al., 2010). In addition, it has more recently been discovered that there is a significant correlation between Offset Vocal Cycle 10 RFF and the perception of vocal effort (Stepp et al., 2012). Thus, RFF may be an indicator of laryngeal tension and consequently may be adapted to serve as an objective marker for vocal hyperfunction. Although RFF may be a promising objective measure, conclusions drawn from examination of RFF in disordered voices are currently limited. Further exploration of RFF in speakers with healthy voices may expedite understanding of the use of RFF-based measures in both healthy and disordered voices.

For instance, examination of the effect of phonetic context in individuals with healthy voices may potentially be used to optimize RFF-based measures applicable for all individuals, including those with disordered voices. Although it has been shown that F0 varies based on context (Fitch, 1990), little research has been done to determine whether the short-term, phonetic variations in F0 captured by RFF are affected by stimuli context. In previous studies, RFF has typically been analyzed from speech sequences extracted from isolated sentences or longer speech passages in which the speech sequences constitute a small proportion of the total stimuli (Goberman & Blomgren, 2008; Robb & Smith, 2002; Stepp, 2013; Stepp et al., 2010, 2011, 2012; Watson, 1998). However, stimuli such as sentences purposefully loaded with voiceless consonants or nonsentence repetitions of speech sequences could potentially also be used to determine RFF. These sentences and utterances are less similar to conversational speech, so they may require an individual to speak more clearly (Ferguson & Kewley-Port, 2002; Payton, Uchanski, & Braida, 1994), providing more reliable utterances for estimation. Furthermore, current estimation of RFF is time-consuming because of the need for trained technicians to manually analyze the recordings. By reducing the syntactic complexity, it may be easier to automate RFF estimation in the future. A possible complication with these types of simplified stimuli is that as utterances shift away from conversational speech, RFF may lose ecological validity. For example, it has been shown that both the mean and overall standard deviation of F0 are higher in reading compared with spontaneous speech for healthy speakers (Horii, 1982; Snidecor, 1943). An examination of the effect of stimulus type on RFF will determine whether stimulus types that require less recording time and are simpler to automate, such as sentences loaded with voiceless consonants or nonsentence repetitions of speech sequences, may be used for RFF estimation.

In addition to stimulus type, RFF may also be affected by the specific voiceless consonant and voiced sonorant. Calculation of RFF centers on the offset and onset of voicing surrounding a voiceless phoneme, but no consistent set of phonemes has been established across studies. Various different phonemes have been used (e.g., /f/, /s/, /k/; Goberman

& Blomgren, 2008; Robb & Smith, 2002; Stepp et al., 2010, 2011; Watson, 1998), potentially drawing into question comparisons across studies. Although Stepp et al. (2010) noted that the choice of voiceless phonemes (/f/ and /k/) used in their study did not have a significant effect on RFF, they examined only three speech sequences (two speech sequences of /f/ in different voiced contexts and one speech sequence of /k/) in eight speakers. An examination of the effect of phoneme identity on RFF may explain the variation in RFF in individuals with similar profiles that were observed across studies that used different voiceless phonemes.

For these reasons, a systematic investigation was carried out to determine the effects of two factors of phonetic context, stimulus type and phoneme identity (/f/, /s/, /ʃ/, /p/, /t/, and /k/), on RFF. To provide linguistically relevant stimuli, we did not control voiced sonorants in these stimuli. Although F0 is known to be dependent on the vowel identity owing to the intrinsic pitch of vowels (Crandall, 1925), the normalization of RFF should minimize the effect of intrinsic pitch; thus, the effect of voiced sonorant was not examined in this study. By determining how the factors of stimulus type and phoneme identity contribute to RFF, the contribution of phonetic context on differences in RFF across studies can be interpreted. In addition, these data will allow for development of a set of stimuli with minimal within-speaker variance for future studies on RFF, resulting in a more reliable objective measure to be applied to future studies in disordered voices.

Method

Participants

Sixteen young adults (eight women; M = 21.5 years, SD = 2.8 years) participated in this study. All participants were native speakers of American English and reported no prior history of speech, language, or hearing disorders. Participants completed written consent in compliance with the Boston University Institutional Review Board.

Speech Stimuli

The speech stimuli consisted of a total of 31 speech tokens: 18 moderately variable and 10 highly variable English sentences, followed by three series of voiced sonorant-voiceless consonant-voiced sonorant, termed *uniform utterances*. The stimuli were produced in the same order by all participants. The uniform utterances consisted of three repetitions of the same speech sequence, /ɑfɑ/, /ifi/, or /ufu/, with each token containing three of the same speech sequence (e.g., /ɑfɑ ɑfɑ ɑfɑ/). The moderately variable sentences were defined as sentences loaded with three speech sequences using the same voiceless consonant phonemes (e.g., "We feel you do fail in new fallen dew" using /f/). Three sentences for each of the following six voiceless consonants were developed: /f/, /s/, /ʃ/, /p/, /t/, and /k/. To ensure consistent elicitations, these consonants were surrounded on both sides by stressed voiced sonorants. In addition, sentences were constructed to ensure that the speech sequences were distinct and that no competing consonants were located in the same sentence. For example, if a speech sequence was targeting the consonant /t/, no other superfluous instances of /t/ were included in the sentence, nor were there instances of its voiced cognate /d/ or any other stop consonants. The highly variable sentences were developed to be loaded with speech sequences containing different voiceless phonemes (e.g., "I saw my five dollar bill in the blue puddle there" using /s/, /f/, and /p/). Four of these sentences contained three speech sequences, and the remaining sentences contained four. All speech stimuli used are listed in Table 1. The degree of phonetic variability is lowest in uniform utterances and highest in highly variable sentences.

Experimental Procedure

Each subject was instructed to read the speech stimuli in his or her comfortable pitch and loudness while wearing a head-mounted microphone (Sennheiser Model PC131) connected to a digital audio recorder (Olympus Model LS-10) recording at 44.1 kHz and 16-bit resolution in a sound-treated room. An experimenter monitored the subject during the task and asked the subject to repeat any stimuli that were misarticulated or obviously glottalized. For the productions of uniform stimuli, the experimenter modeled the utterances before the subject performed the task.

Data Analysis

All recordings were analyzed using Praat acoustic analysis software (Version 5.3.04; Boersma & Weenink, 2012) and Microsoft Excel (Version 14). The default pitch range used in Praat was 60–300 Hz for male recordings and 90–500 Hz for female recordings, although these settings were adjusted for speakers whose pitch fell outside of these ranges. Default settings were used for all other parameters. A single investigator (the first author) computed the periods of the 10 vocal cycles preceding and following the voiceless consonant in each speech sequence. RFF was then calculated by normalizing the instantaneous F0s, the inverse of the periods, relative to reference fundamental frequencies ($F0_{ref}$) in ST using Equation 1:

$$\mathrm{RFF(ST)} = 39.86 \times \log_{10}(\mathrm{F0}/\mathrm{F0}_{ref}) \quad (1)$$

The first offset vocal cycle and 10th onset vocal cycle are closest to the steady-state portions of the offset and onset sonorants, respectively. Thus, the instantaneous F0 for the first offset vocal cycle was chosen as the $F0_{ref}$ for the offset vocal cycles, and instantaneous F0 for the 10th onset vocal cycle was the $F0_{ref}$ for the onset vocal cycles, similar to previous studies (Eadie & Stepp, 2013; Goberman & Blomgren, 2008; Robb & Smith, 2002; Stepp, 2013; Stepp et al., 2010, 2011, 2012; Watson, 1998).

Table 1. Stimuli tokens.

Stimuli	Tokens	Mean Token-Level RFF *SD* (ST)
Moderately variable sentences	We **all fo**und a w**ee fl**y on m**y foo**d on Monday.	0.59
	Nell**y fo**und n**ew fa**bric while R**ay fe**ll down.	0.55
	Only w**e fee**l you d**o fa**il in n**ew fa**llen dew.	0.52
	W**e see** m**y si**bling on m**y si**de mowing.	0.59
	W**e sa**ng a joll**y so**ng all d**ay Su**nday morning.	0.54
	“I saw you b**e si**lly,” Dann**y sa**id angrily.	0.55
	W**e sho**wed Nell m**y shi**ny n**ew shoe** bin.	0.49
	The d**ew shi**mmered over m**y shi**ny bl**ue she**ll again.	0.51
	I w**ish I** would w**ash o**n m**y sho**re one day.	0.61
	I’m h**appy** w**e pa**y our n**ew pa**l Nelly.	0.72
	Lovel**y Pa**mela is y**our pa**l when y**ou pla**y more.	0.65
	The n**ew po**ny loved w**ee Pe**nny and lovel**y Po**lly as well.	0.69
	I tell you, m**y tea** is w**ay too** warm.	0.60
	M**y ti**n**y to**y is a w**ee tra**in with no wheel.	0.69
	I try tearing ever**y to**wel in half.	0.58
	M**y ke**y won h**er ca**r and h**er ca**ne as well.	0.58
	In m**y ca**r y**ou ca**n l**ay ca**lmly.	0.52
	Y**ou kno**ck away m**y ca**ke and Ne**lly ca**me along.	0.72
Highly variable sentences	**I sa**w m**y fi**ve dollar bill in the bl**ue pu**ddle there.	0.54
	Moll**y shi**mmied every even**ing to tu**nes B**o pla**yed her.	0.52
	I called y**ou two** days in a row and y**ou fou**nd a w**ay to** ignore me.	0.63
	A penn**y ca**n only get y**ou so fa**r in life.	0.65
	L**ee sa**w the b**ee fly** in h**er to**p window.	0.62
	M**ay cau**ght the bug with h**er shi**ny bl**ue pa**n bravely.	0.60
	J**oe to**ld her the gr**ay po**ny would tr**y co**ming b**y soo**n again.	0.61
	M**y fa**mil**y sa**w my w**ife** only did whatev**er she** wanted.	0.75
	I said, “**Oh fi**ne, **I’ll sho**w you now.”	0.51
	H**e fu**lly **fe**ll over when h**e sa**w m**y sha**dow there.	0.60
Uniform utterances	/**ɑfɑ ɑfɑ ɑfɑ**/	0.48
	/**ifi ifi ifi**/	0.42
	/**ufu ufu ufu**/	0.47

Note. The speech sequences used to calculate the relative fundamental frequency (RFF) are underlined and bolded. ST = semitones.

Speech sequences were rejected by the investigator if any phonemes were misarticulated, if either voiced sonorants was glottalized, or if the magnitude of the second offset or ninth onset vocal cycle was greater than 0.8 ST. As in previous RFF studies (Stepp, 2013; Stepp et al., 2011, 2012), samples produced with glottalization were excluded because of their irregular vocal cycles. The criteria on the second offset and ninth onset vocal cycles ensured that the reference vocal cycle used for normalization was near steady state. The RFF computed from all three or four speech sequences in a text token (i.e., a series of speech sequences in the uniform stimuli or a sentence in the moderately or highly variable stimuli) were used to calculate the token-level mean and standard deviation used for statistical analysis.

To determine the interrater reliability, a second investigator (the final author) reanalyzed 15% of the samples in each stimulus type (uniform, moderately variable, and highly variable). The Pearson product–moment correlation coefficients were calculated, yielding r = 0.95, 0.93, and 0.94 for uniform utterances, moderately variable sentences, and highly variable sentences, respectively. In addition, the first author reanalyzed 15% of the samples in each stimulus type 3 months after initial analysis to determine the intrarater reliability. The Pearson product–moment correlation coefficients were calculated, yielding r = 0.93, 0.93, and 0.96 for uniform utterances, moderately variable sentences, and highly variable sentences, respectively.

Statistical Analysis

Statistical analysis was performed using Minitab Statistical Software (Version 16.2.2; Minitab, 2010). To determine the effect of stimulus type, a two-factor repeated measures analysis of variance (ANOVA) was performed on the token-level RFF means and standard deviations. Factors were vocal cycle (Offset 1–10 and Onset 1–10), stimulus type, and the interaction between vocal cycle and stimulus type. To determine the effect of phoneme identity, only the moderately variable sentences were used. Again, a two-factor repeated measures ANOVA was performed on the token-level RFF means and standard deviations. The factors were vocal cycle (Offset 1–10 and Onset 1–10), phoneme identity (/f/, /s/, /ʃ/, /p/, /t/, and /k/), and the interaction between vocal cycle and phoneme identity. Effect sizes for each factor were quantified using the squared partial curvilinear correlation (η_p^2) and classified as small, medium, or large (Witte & Witte, 2010). The alpha level for all comparisons was set at .05. All post hoc analyses were completed using Tukey's honestly significant difference tests.

Results

A two-factor repeated measures ANOVA (see Table 2) indicated statistically significant effects of stimulus type (uniform, moderately variable, or highly variable; $p < .001$), vocal cycle ($p < .001$), and the interaction of Vocal Cycle × Stimulus Type ($p < .001$) on the token-level RFF means. The effect sizes of stimulus type and the interaction of Vocal Cycle × Stimulus Type were both small ($\eta_p^2 \leq 0.01$) in comparison to the effect size of vocal cycle ($\eta_p^2 = 0.62$). Post hoc testing revealed that the token-level RFF means elicited from all three different types of stimulus were significantly different from one another. The means of the moderately variable sentences were significantly ($p_{adj} < .05$) higher than those of the highly variable sentences and uniform utterances, and the means of the highly variable sentences were significantly ($p_{adj} < .05$) higher than those of the uniform utterances. To explore these differences in terms of the statistically significant interaction found between stimulus type and vocal cycle, we plotted the token-level RFF means for each type of stimulus as a function of vocal cycle (Figure 2). No substantial difference was observed among the token-level RFF means of each type of stimulus in the offset vocal cycles, except for Offset Vocal Cycle 10, in which the token-level RFF means tended to be slightly lower in highly variable sentences relative to moderately variable sentences and uniform utterances. Minimal differences were observed in the onset vocal cycles but were most pronounced in Onset Vocal Cycles 1–4. In Onset Vocal Cycles 1–4, token-level RFF means tended to be slightly lower in the uniform utterances relative to the highly variable and moderately variable sentences.

Similarly, a two-factor repeated measures ANOVA (see Table 2) indicated statistically significant effects of stimulus type ($p < .001$), vocal cycle ($p < .001$), and the interaction of Vocal Cycle × Stimulus Type ($p = .039$) on the token-level RFF standard deviations. Again, both the effect sizes of stimulus type and the interaction of Vocal Cycle × Stimulus Type were small ($\eta_p^2 = 0.01$) in comparison to the effect of vocal cycle ($\eta_p^2 = 0.45$). Post hoc testing revealed that the token-level RFF standard deviations for highly variable and moderately variable sentences were significantly ($p_{adj} < .05$) higher than those for uniform utterances, but no statistically significant ($p_{adj} < .05$) difference in token-level RFF standard deviations was found between highly variable and moderately variable sentences. Figure 3 shows a plot of the mean values of the token-level RFF standard deviations in each stimuli category. Examination of the token-level RFF standard deviations as a function of vocal cycle in each category (see Figure 4) revealed that the means for uniform utterances were lowest for all nonreference vocal cycles, but the difference was most pronounced in Onset Vocal Cycles 1–2.

Figure 2. Token-level RFF means as a function of stimulus type (uniform: uniform utterances; moderate: moderately variable sentences; and high: highly variable sentences) and vocal cycle (Offset 1–10 and Onset 1–10) in semitones (ST). Uniform utterances consisted of three repetitions of the same speech sequences. Moderately variable sentences contained speech sequences of a single voiceless phoneme. Highly variable sentences were loaded with speech sequences using multiple phonemes. Error bars indicate the 95% confidence intervals for the means.

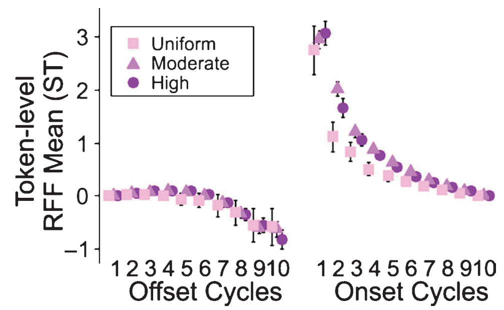
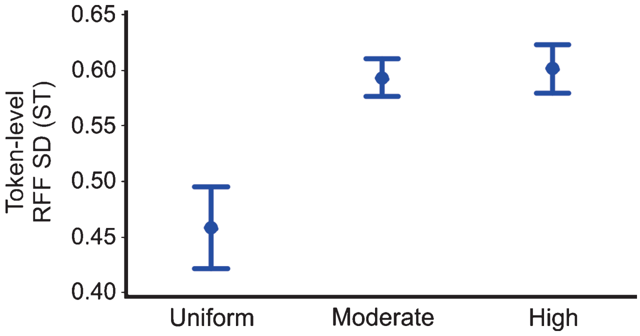


Table 2. Results of two-factor repeated measures analysis of variance (ANOVA) on token-level RFF means and standard deviations.

Effect	*df*	η_p^2	*F*	*p*
Two-factor (vocal cycle, stimulus type) repeated measures ANOVA				
Vocal cycle (Offset 1–10, Onset 1–10)	19	*M* = 0.62	*M* = 467.6	*M* < .001
		SD = 0.45	SD = 233.0	SD < .001
Stimulus type (uniform, moderate, high)	2	*M* < 0.01	*M* = 22.2	*M* < .001
		SD = 0.01	SD = 39.0	SD < .001
Vocal Cycle × Stimulus Type	38	*M* = 0.01	*M* = 3.2	*M* < .001
		SD = 0.01	SD = 1.4	SD = 0.039
Two-factor (vocal cycle, phoneme identity) repeated measures ANOVA				
Vocal cycle	19	*M* = 0.64	*M* = 518.4	*M* < .001
		SD = 0.47	SD = 252.0	SD < .001
Phoneme identity (/f/, /s/, /ʃ/, /p/, /t/, /k/)	5	*M* = 0.03	*M* = 32.4	*M* < .001
		SD = 0.01	SD = 13.2	SD < .001
Vocal Cycle × Phoneme Identity	95	*M* = 0.07	*M* = 4.4	*M* < .001
		SD = 0.02	SD = 1.4	SD = 0.012

Note. The lower portion of the table shows means and standard deviations for moderately variable sentences only. Uniform = uniform utterances; Moderate = moderately variable sentences; High = highly variable sentences.

Figure 3. Token-level RFF standard deviations as a function of stimulus type (uniform, moderate, and high) in ST. Error bars indicate the 95% confidence intervals for the means.



A two-factor repeated measures ANOVA on token-level RFF means (see Table 2), revealed statistically significant effects of phoneme identity ($p < .001$), vocal cycle ($p < .001$), and the interaction of Vocal Cycle × Phoneme Identity ($p < .001$). There was a small effect size for phoneme identity ($\eta_p^2 = 0.03$), a medium effect size for the interaction of Vocal Cycle × Phoneme Identity ($\eta_p^2 = 0.07$), and a large effect size for vocal cycle ($\eta_p^2 = 0.64$). Post hoc testing revealed that the RFF means for both /f/ and /ʃ/ were significantly ($p_{adj} < .05$) higher than for /s/, /p/, /t/, and /k/ and that the RFF means for /s/, /p/, and /t/ were significantly lower than for /k/. To further explore these relationships, we plotted the token-level RFF means as a function of vocal cycle for each phoneme (Figure 5). Compared with stops (/p/, /t/, /k/), fricatives (/f/, /s/, and /ʃ/) tended to show higher RFF for the vocal cycles nearest to the consonant (i.e., Offset Vocal Cycle 10 and Onset Vocal Cycle 1).

A two-factor repeated measures ANOVA on token-level RFF standard deviations (see Table 2) revealed statistically significant effects of phoneme identity ($p < .001$), vocal cycle ($p < .001$), and interaction of Vocal Cycle × Phoneme Identity ($p = .012$). There was a small effect size for phoneme identity ($\eta_p^2 = 0.01$) and the interaction of Vocal Cycle × Phoneme Identity ($\eta_p^2 = 0.02$) and a large effect size for vocal cycle ($\eta_p^2 = 0.47$). Post hoc testing revealed several significant ($p_{adj} < .05$) differences between the phonemes. A plot of the mean values of token-level standard deviations for each phoneme is shown in Figure 6. The mean standard deviation for /p/ was the highest, whereas the mean standard deviations for the fricatives (/f/, /s/, and /ʃ/) were the lowest. Examination of the mean values of token-level standard deviations as a function of vocal cycle revealed that fricatives tended to have lower or equal standard deviations compared with stops for all nonreference vocal cycles, except for Onset Vocal Cycles 1–2 in which marginal differences were observed between fricatives and stops.

Figure 4. Token-level RFF standard deviations as a function of stimulus type (uniform, moderate, and high) and vocal cycle (Offset 1–10 and Onset 1–10) in ST.

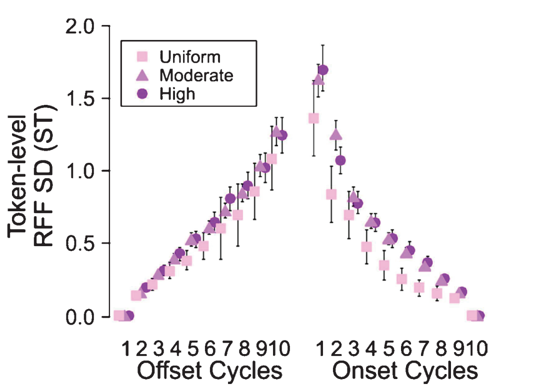
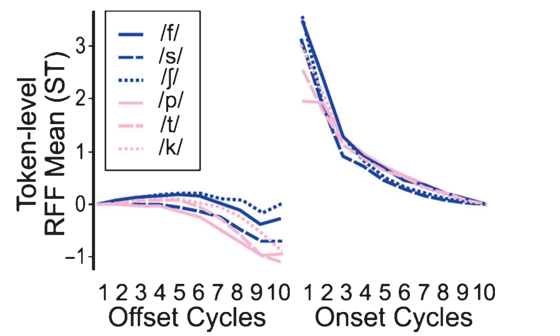


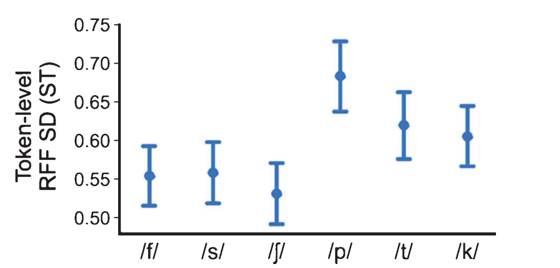
Figure 5. Token-level RFF means as a function of phoneme identity (/f/, /s/, /ʃ/, /p/, /t/, and /k/) and vocal cycle (Offset 1–10 and Onset 1–10) in ST. Fricatives (/f/, /s/, and /ʃ/) are plotted in dark blue, whereas stop consonants (/p/, /t/, and /k/) are plotted in light pink. Error bars are omitted for clarity.



Discussion

The goal of this study was to understand how phonetic context affects RFF. Statistically significant effects of stimulus type and phoneme identity were found for token-level RFF means; however, the effect sizes for stimuli type and phoneme identity were generally small. When compared across stimulus type, RFF means were highest in moderately

Figure 6. Token-level RFF standard deviations as a function of phoneme identity (/f/, /s/, /ʃ/, /p/, /t/, and /k/) in ST. Error bars indicate the 95% confidence intervals for the means.



variable sentences and lowest in uniform utterances. When compared across phoneme identity, RFF means were highest in sentences using the phonemes /f/ and /ʃ/ and lowest in those containing /p/.

Similarly, statistically significant effects of stimulus type and phoneme identity on token-level RFF standard deviations were found, but again the effect sizes were small. When compared across stimulus type, RFF standard deviations were similar between the highly and moderately variable sentences and lowest in uniform utterances. When compared across phoneme identity, RFF standard deviations were highest in speech sequences containing /p/ and lowest in those containing /f/ and /ʃ/.

Effects of Stimulus Type

The results of the study indicate that differences in mean RFF were most noticeable in Onset Vocal Cycles 1–4, with the lowest means in the uniform utterances (see Figure 2). Minimal differences were seen in the offset vocal cycles, with the exception of Offset Vocal Cycle 10, in which the token-level RFF means for the highly variable stimulus tended to be slightly lower than those for moderately variable and uniform stimuli. The effect of stimulus type on mean RFF was primarily on the onset vocal cycles, unlike the effects of gender (Stepp, 2013), aging (Watson, 1998), and voice disorders (Stepp et al., 2010, 2012), which were largely on offset vocal cycles.

In addition, stimulus type showed a small but significant effect on RFF standard deviations. Compared with both highly variable and moderately variable sentences, uniform utterances elicited RFF with the lowest token-level standard deviations (see Figure 3). One potential reason for this difference may be the difference in the amount of variation in the flanking sonorants in the three stimulus types. In the uniform utterances, the RFF standard deviations were computed for speech sequences with exactly the same flanking sonorants throughout the utterance (e.g., /ɑfɑ/ /ɑfɑ/ /ɑfɑ/; all /ɑ/), but in highly variable and moderately variable sentences, they were computed for speech sequences with different flanking sonorants to allow for linguistically appropriate stimuli. Although the effect of sonorants on RFF has not been studied previously, averaging RFF over different voiced sonorants potentially could lead to higher variance because of differences in the intrinsic pitch of vowels (Crandall, 1925). Conversely, the instantaneous F0s of each sonorant are normalized to the $F0_{ref}$ of that same sonorant during the calculation of RFF, so the effect of intrinsic pitch should be reduced or diminished. Nevertheless, future studies are warranted to systematically determine the effect of surrounding sonorants on token-level RFF standard deviations.

Another potential hypothesis for why uniform utterances tended to have lower standard deviations compared with highly variable and moderately variable sentences is that the duration of sonorants for speech sequences in sentences may be shorter than in nonlinguistic stimuli, resulting in reference vocal cycles that are not truly at steady state. Token-level standard deviations should be higher for sentences if some of the speech sequences are poorly normalized. To explore this hypothesis, we compared the absolute values of Offset Vocal Cycle 2 and Onset Vocal Cycle 9 (vocal cycles adjacent to the reference vocal cycles) in the three types of stimuli. We expected the absolute RFF values of these vocal cycles to be small, near zero ST if the reference vocal cycles have reached steady state. When the mean absolute RFF in the vocal cycles adjacent to the reference vocal cycles was calculated, uniform utterances had somewhat lower values (offset mean = 0.083 ST; onset mean = 0.077 ST) when compared with moderately variable sentences (offset mean = 0.11 ST; onset mean = 0.12 ST) and highly variable sentences (offset mean = 0.097 ST; onset mean = 0.12 ST). This post hoc analysis supports the hypothesis that the reference vocal cycles in speech sequences in uniform utterances may have been closer to steady state compared with those in moderately variable and highly variable sentences, thus contributing to the small effect of stimulus type noted for RFF standard deviations.

Effects of Phoneme Identity

The effects of phoneme identity and the interaction of Vocal Cycle × Phoneme Identity were statistically significant with small and medium effect sizes, respectively. The most obvious differences in RFF between phonemes were observed in Offset Vocal Cycles 4–10 and Onset Vocal Cycle 1, in which RFF tended to be higher in fricatives than in stops (see Figure 5). These vocal cycles affected by phoneme identity are similar to those that have been shown to be affected by vocal hyperfunction (Stepp et al., 2010). Thus, it may be especially important to use stimuli with consistent phonemes when using RFF to evaluate vocal hyperfunction.

Although the underlying physiological differences that result in the differences in mean RFF observed in stops and fricatives are not known, the differences may result from differences in vocal tract activity associated with these two different manners of production. During the production of both voiceless stops and fricatives, the vocal folds are abducted to allow pressure to build in the oral cavity of the vocal tract. However, the production of stops requires full oral constriction followed by a release, whereas fricatives require partial oral constriction throughout the entire duration of the consonant. Fricatives are expected, in comparison to stops, to have higher airflow preceding the voiceless consonant so that the frication can start immediately following the vowel and to have lower airflow following the voiceless consonant because less pressure has built up during the partial constriction relative to the full constriction in stops. In fact, Löfqvist et al. (1995) measured the mean minimum and mean peak airflows in vowel–consonant–vowel utterances produced by two subjects and found that airflow was higher preceding and lower following the production of the fricative /s/ compared with the production of the stop /p/. Thus, the higher offset RFF observed in fricatives may be a result of this increased airflow. However, the lower airflow following the fricative should lead to lower onset RFF for fricatives relative to stops, which is contrary to the observations in this study.

Another potential explanation for differences in onset RFF between stops and fricatives is due to differences in the laryngeal activity. Notably, the laryngeal tension in the vocal folds preceding and following the obstruents may differ in the production of fricatives relative to stops. Löfqvist et al. (1989) measured the activity of the cricothyroid muscle during both voiceless stops and fricatives. Although the purpose of their study was not to compare these two types of obstruents, they found that the mean muscle activity measured during stops was lower or equal to that during the fricatives for all three subjects studied. In another study (Fant, 1970), the activity of the vocalis muscle was measured from a single Dutch speaker. Results indicated that the relaxation of the vocalis muscle was greater when the speaker transitioned from the fricative /f/ to /ɑ/ than when the speaker transitioned from the stop /t/ to /ɑ/. These studies suggest that more tension may be used for revoicing following fricatives compared with stops. This difference in tension might lead to higher onset RFF in fricatives, which we observed in our study. However, because of differences in the intent of study and limited sample sizes, neither study provides conclusive evidence that there is a difference in tension in voiced sonorants preceding or following stops relative to fricatives or that this difference drives differences in offset or onset RFF. Future studies involving simultaneous physiological measurements of tension, airflow, and vocal-fold kinematics during the production of RFF stimuli should be conducted to determine the underlying physiology behind these differences.

Implications and Limitations

A set of stimuli with minimal intraspeaker variability can be developed on the basis of results of the study. Given that the lowest token-level RFF standard deviations were seen for uniform stimulus and the phonemes /f/ and /ʃ/ and that the expected patterns of phonemic-level changes in F0 were still preserved in the uniform utterances, uniform utterances with the phonemes /f/ and /ʃ/ are recommended as stimuli for future studies and clinical protocols.

The results of this study should be interpreted with respect to several limitations. First, the finding of an effect of stimulus type on RFF standard deviation should be interpreted with caution. Although results showed that uniform utterances had the lowest token-level standard deviation, the uniform utterances tested were all speech sequences with the phoneme /f/, and this particular phoneme was found to have one of the lowest token-level RFF standard deviations. Second, the stimuli were always produced in the order of moderately variable sentences followed by highly variable sentences and uniform utterances, which could potentially have resulted in an order effect. Participants may have become more experienced or fatigued as the experiment progressed. Previous studies have shown that F0 tends to increase after prolonged voice use (Lehto, Laaksonen, Vilkman, & Alku, 2008; Stemple, Stanley, & Lee, 1995). However, the duration of the experiment in this study generally took less than 30 min, so the difference in baseline F0 was likely minimal, and the calculation of RFF takes into account the baseline F0 by normalizing instantaneous F0s to a steady-state F0. Third, although all participants were speakers of American English, individual dialects were not recorded and could have had effects on pronunciation and thus RFF. Finally, the current results can be generalized only to young speakers with healthy voices. Future studies involving older adults and speakers with disordered voices should be carried out to determine whether stimulus type and phoneme identity have similar effects in these speakers. Perceptual studies should also be carried out to determine how well RFF estimated using different types of stimuli correlates with the perception of voice quality. Eadie and Stepp (2013) showed that individuals with adductor spasmodic dysphonia need at least six speech sequences to achieve a stable RFF estimate that correlates with listener perceptions of vocal effort. However, these stimuli were taken from a single sentence loaded with nine speech sequences with a variety of different phonemes. Fewer speech sequences may be needed for a stable RFF estimate that correlates with perception if stimuli with inherently lower RFF standard deviations are used.

A significant effect of the interaction between vocal cycle and phoneme identity on mean RFF with a medium effect size was found. Closer inspection indicated differences in both offset and onset RFF as a function of the manner of production (stops vs. fricatives). This finding leaves questions about the differences in physiology and RFF production in stops and fricatives. Future studies involving a neck-placed miniature accelerometer (Hillman, Heaton, Masaki, Zeitels, & Cheyne, 2006) could be used to determine vocal-fold vibrations that might be masked by the sudden high energy releases in stops and fricatives in the acoustic signal. In addition, simultaneous measures of oral airflow, electromyography of muscles involved in laryngeal tension (suprahyoid, cricothyroid, and vocalis muscles), and imaging should be used to determine the contributions of tension, aerodynamics, and vocal-fold kinematics to RFF production.

Conclusion

Stimulus type and phoneme identity have mostly small but significant effects on token-level RFF means and standard deviations. Thus, it may be necessary to account for the effect of phonetic context when comparing RFF across studies. To minimize within-speaker RFF variability, uniform utterances with the phonemes /f/ and /ʃ/ are recommended because they showed the lowest token-level RFF standard deviations. Future studies are necessary to determine the effects of stimulus type and phoneme identity on RFF in speakers with disordered voices.

Acknowledgments

This work was supported by the National Institute on Deafness and Other Communication Disorders Grant DC012651 and a New Century Scholars grant from the American Speech-Language-Hearing Foundation.

References

Arnold, G. E. (1961). Physiology and pathology of the cricothyroid muscle. *Laryngoscope, 71,* 687–753. doi:10.1288/00005537-196107000-00002

Berardelli, A., Sabra, A. F., & Hallett, M. (1983). Physiological mechanisms of rigidity in Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry, 46,* 45–53.

Boersma, W., & Weenink, D. (2012). Praat: Doing phonetics by computer (Version 5.3.04) Computer program. Retrieved from http://www.praat.org/

Crandall, I. B. (1925). The sounds of speech. *Bell System Technical Journal, 4,* 586–626.

Eadie, T. L., & Stepp, C. E. (2013). Acoustic correlate of vocal effort in spasmodic dysphonia. *Annals of Otology, Rhinology, and Laryngology, 122,* 169–176.

Fant, G. (1970). *Acoustic theory of speech production* (2nd ed.). The Hague, the Netherlands: Mouton.

Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 112,* 259–271.

Fitch, J. L. (1990). Consistency of fundamental frequency and perturbation in repeated phonations of sustained vowels, reading, and connected speech. *Journal of Speech and Hearing Disorders, 55,* 360–363.

Fukui, N., & Hirose, H. (1983). Laryngeal adjustments in Danish voiceless obstruent production. *Annual Report of the Institute of Phonetics, University of Copenhagen, 17,* 61–71.

Gallena, S., Smith, P. J., Zeffiro, T., & Ludlow, C. L. (2001). Effects of levodopa on laryngeal muscle activity for voice onset and offset in Parkinson disease. *Journal of Speech, Language, and Hearing Research, 44,* 1284–1299.

Goberman, A. M., & Blomgren, M. (2008). Fundamental frequency change during offset and onset of voicing in individuals with Parkinson disease. *Journal of Voice, 22,* 178–191.

Hillman, R. E., Heaton, J. T., Masaki, A., Zeitels, S. M., & Cheyne, H. A. (2006). Ambulatory monitoring of disordered voices. *Annals of Otology, Rhinology, and Laryngology, 115,* 795–801.

Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech and Hearing Research, 32,* 373–392.

Horii, Y. (1982). Some voice fundamental frequency characteristics of oral reading and spontaneous speech by hard-of-hearing young women. *Journal of Speech and Hearing Research, 25,* 608–610.

Ladefoged, P. (1972). *Three areas of experimental phonetics: Stress and respiratory activity, the nature of vowel quality, units in the perception and production of speech* (Vol. 15). London, England: Oxford University Press.

Lehto, L., Laaksonen, L., Vilkman, E., & Alku, P. (2008). Changes in objective acoustic measurements and subjective voice complaints in call center customer-service advisors during one working day. *Journal of Voice, 22,* 164–177.

Löfqvist, A., Baer, T., McGarr, N. S., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *The Journal of the Acoustical Society of America, 85,* 1314–1321.

Löfqvist, A., Koenig, L. L., & Mcgowan, R. S. (1995). Vocal-tract aerodynamics in /aca/ utterances: Measurements. *Speech Communication, 16,* 49–66.

Löfqvist, A., & Mcgowan, R. S. (1992). Influence of consonantal environment on voice source aerodynamics. *Journal of Phonetics, 20,* 93–110.

Minitab. (2010). Minitab Statistical Software (Version 16.2.2) [Computer program]. State College: Pennsylvania State University.

Payton, K. L., Uchanski, R. M., & Braida, L. D. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America, 95,* 1581–1592.

Robb, M. P., & Smith, A. B. (2002). Fundamental frequency onset and offset behavior: A comparative study of children and adults. *Journal of Speech, Language, and Hearing Research, 45,* 446–456.

Roubeau, B., Chevrie-Muller, C., & Lacau Saint Guily, J. (1997). Electromyographic activity of strap and cricothyroid muscles in pitch change. *Acta Oto-Laryngologica, 117,* 459–464.

Roy, N., Ford, C. N., & Bless, D. M. (1996). Muscle tension dysphonia and spasmodic dysphonia: The role of manual laryngeal tension reduction in diagnosis and management. *Annals of Otology, Rhinology, and Laryngology, 105,* 851–856.

Snidecor, J. C. (1943). A comparative study of the pitch and duration characteristics of impromptu speaking and oral reading. *Communications Monographs, 10,* 50–56.

Stemple, J. C., Stanley, J., & Lee, L. (1995). Objective measures of voice production in normal subjects following prolonged voice use. *Journal of Voice, 9,* 127–133.

Stepp, C. E. (2013). Relative fundamental frequency during vocal onset and offset in older speakers with and without Parkinson's disease. *The Journal of the Acoustical Society of America, 133,* 1637–1643.

Stepp, C. E., Hillman, R. E., & Heaton, J. T. (2010). The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research, 53,* 1220–1226.

Stepp, C. E., Merchant, G. R., Heaton, J. T., & Hillman, R. E. (2011). Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 54,* 1260–1266.

Stepp, C. E., Sawin, D. E., & Eadie, T. L. (2012). The relationship between perception of vocal effort and relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research, 55,* 1887–1896.

Watson, B. C. (1998). Fundamental frequency during phonetically governed devoicing in normal young and aged speakers. *The Journal of the Acoustical Society of America, 103,* 3642–3647.

Witte, R. S., & Witte, J. S. (2010). *Statistics* (9th ed.). Hoboken, NJ: Wiley.