Empirical Evaluation of the Role of Vocal Fold Collision on Relative Fundamental Frequency in Voicing Offset

[*,†]**Matti D. Groll**, [‡]**Sean D. Peterson**, [§]**Matías Zañartu**, [*,†]**Jennifer M. Vojtech, and** [*,†,‖]**Cara E. Stepp**, *†‖ *Boston, Massachusetts,* ‡*Ontario, Canada, and* §*Valparaíso, Chile*

Summary: Objectives. Relative fundamental frequency (RFF) is an acoustic measure of changes in fundamental frequency during voicing transitions. The physiological mechanisms underlying RFF remain unclear. Recent modeling suggests that changes in RFF during voicing offset are due to decreases in overall system stiffness as a direct result of the cessation of vocal fold collision. To evaluate this finding empirically, here we examined whether variable timing between the end of vocal fold collision and the final voicing cycle used to calculate RFF explained the variability in RFF across individual voicing offset utterances.

Methods. RFF during voicing offset was calculated from /ifi/ utterances produced by 35 participants under endoscopy, with and without vocal effort. RFF was calculated via two methods, in which utterances were aligned by (1) the end of vocal fold collision, or (2) the end of voicing. Analyses of variance were used to determine the effects of vocal effort and RFF method on the mean and standard deviation of RFF.

Results. Aligning by vocal fold collision resulted in statistically significantly lower standard deviations. RFF means were statistically higher using the collision method; however, the degree of vocal effort was statistically significant regardless of the method.

Conclusions. These results provide empirical evidence to support that decreases in RFF during voicing offset are a result of decreases in system stiffness due to termination of vocal fold collision.

Key Words: Laryngeal tension—Relative fundamental frequency—Vocal fold kinematics.

INTRODUCTION

Relative fundamental frequency (RFF) is an acoustic measure that calculates changes in the cycle-by-cycle fundamental frequency (f_o) during voicing transitions.[1] Typically, RFF is calculated during a vowel-voiceless consonant-vowel utterance eg,[1-3] such that the ten voiced cycles immediately preceding the voiceless consonant are RFF offset and the ten voiced cycles immediately following the voiceless consonant are RFF onset. RFF is calculated in semitones (ST) in relation to a speaker's steady-state f_o using Equation 1, in which f_o^{ref} is the f_o of the relatively steady-state reference cycle (corresponding to the first cycle or the tenth cycle for offset and onset, respectively), and $f_o^{\,i}$ is the f_o of cycle i. When averaged across multiple utterances, a typical RFF pattern emerges (Figure 1). Specifically, RFF offset begins at zero ST and tends to decrease for cycles closer to the voiceless consonant, particularly in older adults.[4] In contrast, RFF onset values are positive for cycles closer to the voiceless consonant and then trend toward zero.[5-9] This pattern has been shown to be relatively consistent across individuals with typical voices in many previous studies.[1-3,10-17]

$$RFF^i\ (\mathrm{ST}) = 39.86 \times \log_{10}\left(\frac{f_o^i}{f_o^{\mathrm{ref}}}\right) \tag{1}$$

Previous work has suggested that RFF measures may be sensitive to changes in baseline levels of laryngeal tension due to increases in laryngeal muscle activation. Specifically, when estimates of laryngeal tension increase, RFF decreases, particularly in cycles nearest the voiceless consonant [ie, offset cycle 10 and onset cycle 1; 2]. When compared to speakers with typical voices, speakers with voice disorders that are characterized by increased laryngeal tension, such as vocal hyperfunction,[1,18,19] laryngeal dystonia,[20] and Parkinson's disease,[12,15] have all demonstrated decreased RFF. A previous study also found that, following successful voice therapy, individuals with vocal hyperfunction showed increases in RFF that trended toward the RFF values of individuals with typical voices.[13] Furthermore, when speakers with typical voices are instructed to speak with increased levels of vocal effort, RFF decreases.[2,3] Vocal effort is defined as an individual's self-perception of the level of exertion needed to produce a response to a given communication scenario.[21] Increased levels of vocal effort are often observed in individuals with voice disorders characterized by increases in laryngeal tension,[18] suggesting that vocal effort may be associated with increases in laryngeal tension.[2,22] This sensitivity to changes in estimates of laryngeal tension makes RFF a potential acoustic measure for the assessment of voice disorders.

Despite an association between RFF and laryngeal tension, the physiological mechanisms behind RFF remain relatively unknown. It has been proposed that changes in

Accepted for publication September 19, 2022.

Supported by the National Institutes of Health grants R01 DC015570, P50 DC015446, and T32 DC013017 from the National Institute on Deafness and Other Communication Disorders (NIDCD), as well as ANID grant BASAL FB0008.

From the *Department of Biomedical Engineering, Boston University, Boston, Massachusetts; †Department of Speech, Language, and Hearing Sciences, Boston University, Boston, Massachusetts; ‡Department of Mechanical and Mechatronics Engineering, University of Waterloo, Ontario, Canada; §Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile; and the ‖Department of Otolaryngology-Head and Neck Surgery, Boston University School of Medicine, Boston, Massachusetts.

Address correspondence and reprint requests to Matti D. Groll, Department of Biomedical Engineering, Boston University, 677 Beacon Street, Boston, MA, 02215. E-mail: mgroll@bu.edu

Journal of Voice, Vol. ■■, No. ■■, pp. ■■−■■

0892-1997

https://doi.org/10.1016/j.jvoice.2022.09.016

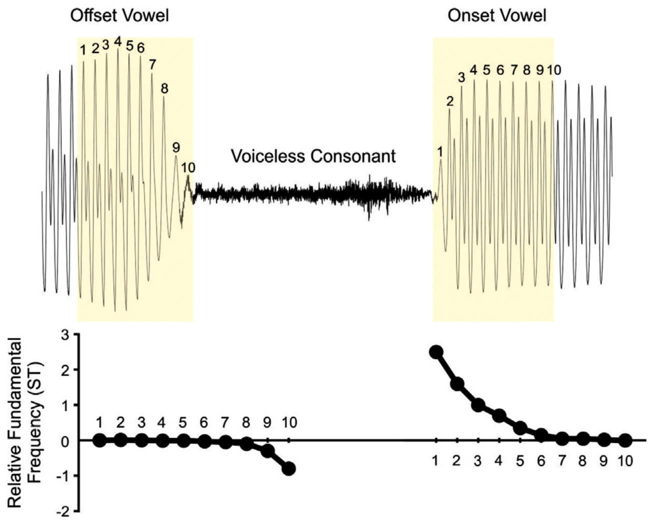

FIGURE 1. Example of a typical relative fundamental frequency (RFF) pattern for ten offset and ten onset cycles. RFF is calculated using Equation 1. Offset cycle 1 and onset cycle 10 are the reference cycles for RFF offset and onset, respectively.

cycle-by-cycle f_o during voicing transitions are due to a combination of changes in transient tension, vocal fold abduction, and aerodynamic forces.[4,13,23-28] Laryngeal tension is thought to temporarily increase throughout the production of a voiceless consonant, causing an increase in f_o immediately prior to and following the voiceless consonant.[23] In contrast, the abduction of the vocal folds during voicing offset is thought to decrease f_o.[4] In theory, then, this combination of increased tension and abduction results in a small net decrease in RFF offset, as shown in Figure 1. During the release of the voiceless consonant and the subsequent onset of voicing, it has been hypothesized that high rates of airflow cause rapid adduction and increases in f_o.[29] Thus, the combination of increased tension and increased airflow contribute to the large increase in RFF during voicing onset as shown in Figure 1.

This theoretical framework for the physiological mechanisms behind RFF may also be used to explain decreases in RFF due to increases in baseline levels of laryngeal tension. In a study that found that RFF in speakers with vocal hyperfunction is lower than in speakers with typical voices, the authors reasoned that the individuals with vocal hyperfunction had a restricted ability to further increase the level of tension in the laryngeal muscles during the offset and onset of voicing.[1] Thus, when evaluating the combined interactions between tension, abduction, and aerodynamic forces, there would be a smaller increase in f_o due to the baseline tension, thus lowering both RFF offset and onset.

Although this physiological framework may explain typical RFF offset and onset patterns, these patterns are most often observed after averaging RFF values across multiple utterances. Yet RFF may vary within a speaker substantially. A previous study recommended that an average of six utterances be used for stable RFF estimates.[20] This within-speaker variability limits the clinical feasibility of RFF as a reliable acoustic measure of laryngeal tension. Thus, it is important to understand which factors contribute to RFF variability across individual utterances. Although previous studies have shown that phonemic context and stress type have an effect on RFF and should be consistent during recordings,[11,16,27,30] even identical utterances produced by the same speaker in succession may demonstrate variability.

Recently, modeling has been used to further investigate the physiological mechanisms behind RFF, which may help explain the variability across phonetically identical utterances. Serry, Stepp, and Peterson[31] employed two models to examine the physics behind phonation offset. They first used a simple impact oscillator model as a proxy for the vocal folds to demonstrate that the system f_o was higher when the model folds were colliding with one another than when they were sufficiently abducted such that collision no longer occurred. Collision of the impact oscillator was modeled as an additional spring stiffness for the masses in contact, such that, during the collision regime, there was a greater overall stiffness of the vocal folds than during the non-collision regime, when system stiffness comprised only of the vibrating vocal fold material.[31] Thus, as the vocal folds abduct during voicing offset and transition from the collision regime to the non-collision regime, there is a transient decrease in the stiffness of the system, which causes a decrease in f_o.

Following the simple impact oscillator model, Serry, Stepp, and Peterson[31] then performed a numerical investigation using the body-cover model,[32] wherein the vocal folds were modeled as three masses connected via springs and dampers to model tissue viscoelasticity: two impact oscillator masses served as the vocal fold cover and were connected to a third body mass. This numerical work further showed that decreases in RFF were correlated with the transient drop in collision forces following the end of vocal fold contact.[31]

If the timing between the end of vocal fold contact and the final voicing cycle used in RFF measures varies across utterances in a single speaker, this may partially explain the documented within-speaker RFF variability. Though previous studies rarely report intraspeaker timing variability across identical utterances, one previous study showed that the time between the end of vocal fold contact and the end of voicing varied within speaker across tokens of /ifi/ and /iti/ by as much as 40 ms, corresponding to several voicing cycles depending on the f_o of the speaker.[33] With a variable time between the end of vocal fold contact and a given offset cycle, the decrease in overall system stiffness resulting from transitioning from the collision regime to the non-collision regime would have a variable amount of impact on RFF from utterance to utterance. For example, RFF cycle 10 of one utterance may correspond to the cycle immediately following the end of vocal fold contact, whereas RFF cycle 10 of another utterance may correspond to the fifth cycle following the end of vocal fold contact. If the relationship between the timing of the end of vocal fold contact and the decrease in RFF offset is valid, then comparing RFF cycle 10 values between these two utterances would likely result in large variability. Furthermore, if RFF was calculated by aligning offset cycles based on a set duration away from the end of vocal fold contact instead of by aligning offset cycles based on the final phonation cycle, and a subsequent decrease in the within-speaker variability of RFF occurred, this would provide empirical evidence supporting the relationship between the end of vocal fold contact and RFF offset as proposed by these models.[31]

The purpose of this study was to investigate and validate the theoretical relationship between the end of vocal fold contact and RFF offset by answering two research questions. First, how does aligning RFF offset based on a set duration away from the last point of vocal fold contact affect within-speaker variability in the final RFF cycle when compared to traditional RFF that is always aligned based on the end of voicing, irrespective of when vocal fold contact ceased? Based on the framework of the vocal fold models employed in,[31] we hypothesized that aligning RFF offset based on the last point of vocal fold contact instead of the end of phonation would result in decreased within-speaker variability in the final offset RFF cycle as a result of a more consistent effect from the transient decrease in system stiffness across all utterances. Second, does this novel method of calculating RFF show the same sensitivity to changes in vocal effort as traditional RFF has shown in previous studies? This research question was used to confirm the validity of the novel RFF method. We hypothesized that both methods of calculating RFF would demonstrate a similar decrease in RFF when individuals were instructed to speak with increased levels of vocal effort.

METHODS

Participants

A total of 45 young adult speakers with typical voices (24 female, 21 male; M = 21.9 years, SD = 3.5 years) were recruited for this study. Twenty of the 45 participants were cisgender. Gender data were not available for the other 25 participants. All participants were native speakers of American English, were non-smokers, reported no history of voice disorders or laryngeal abnormalities, and passed a hearing screening indicating typical hearing. The hearing screening presented monaural pulsed tones at a range of frequencies between 125 Hz to 8000 Hz. Participants were considered to have typical hearing if all tones could be detected at 25 dB HL. All individuals completed written consent under the guidelines of the Boston University Institutional Review Board.

Data collection

All participants were seated in a sound-treated booth for the duration of the experiment. A directional headset microphone was placed on each participant, 45° from the midline of the vermillion and 7 cm from the corner of the lips. Participants were instructed to produce a series of vowel-voiceless consonant-vowel (/ifi/) utterances under endoscopy, performed by a certified speech-language pathologist. These /ifi/ utterances were produced with no vocal effort and with maximum vocal effort. For each recording, participants were trained using a metronome (65 beats per minute) to produce seven to eight /ifi/ utterances at a time. Recordings at each effort condition were performed twice, for an approximate total of 14 to 16 /ifi/ utterances per effort condition. For no vocal effort recordings, participants were instructed to speak with their typical voice. For effort recordings, participants were given the following cue: "Now we would like you to increase your effort during your speech as if you are trying to create tension in your voice as if you are trying to push your air out. Try to maintain the same volume while increasing your effort." Participants were also instructed to maintain a comfortable speaking rate and pitch, as well as a typical vocal volume across all recordings. The maximum vocal effort condition was defined as "as much effort as you can, while still maintaining a voice." Participants practiced producing utterances with the experimenter in order to verify appropriate rate and vocal effort, as well as consistent speaking rate, pitch, and volume for each condition.

During the production of /ifi/ utterances, endoscopic video data were collected via a flexible nasal endoscope (Pentax, Model FNL-10RP3, 3.5-mm) by a certified speech-language pathologist. If the participant expressed discomfort during the initial insertion through the nasal passages, a pediatric nasal endoscope was used instead (Pentax, Model FNL-7RP3, 2.4-mm). Endoscopic video images of the vocal folds were recorded at a frame rate of 1 kHz, which has been shown to be a suitable frame rate for capturing abductory kinematics while maintaining an adequate level of lighting.[34] The endoscope was attached to a camera (FASTCAM Mini AX100l; Model 540K-C-16GB; 256 × 256 pixels) with a 40-mm optical lens adapter. Constant xenon light was used for imaging (300 W KayPentax Model 7162B) and video images were acquired using Photron Fastcam Viewer software (v.3.6.6). Recordings were triggered using a custom MATLAB algorithm to automatically time-align the video images with the microphone signals.

Videoendoscopic data analysis

Videoendoscopic data of the vocal folds were used to determine timing parameters for the offset and subsequent onset of voicing for each /ifi/ utterance. Only voicing offset was considered in the current study. Three trained technicians manually inspected videoendoscopic data to estimate the time during vocal fold abduction when the vocal folds ceased to collide (ie, "last point of contact"). For utterances in which technicians were unable to identify this time point (eg, if the glottis was obstructed or the endoscopic image was too dark or blurry), the utterance was omitted from further analysis (see Utterance Rejection Criteria).

Each of the three technicians were trained to mark the time point corresponding cessation of vocal fold collision by performing markings on the small set of videoendoscopic data comprised of /ifi/ utterances from three speakers not included in the current study. Interrater reliability for vocal fold contact marking was calculated between technicians using a two-way intraclass correlation (ICC) analysis. The average interrater reliability was determined to be ICC (2,1) = 0.85. Technicians repeated markings on this training set in a separate session to determine intrarater reliability, which was calculated as ICC (2,1) = 0.99, 0.98, and 0.99 for the three raters. These results indicate good interrater reliability and excellent intrarater reliability.[35]

RFF methods

Semi-automated RFF algorithms[36] were adapted to automatically calculate the locations of individual voicing cycles during voicing offset. The final voicing cycle identified by the RFF algorithms was considered to be the time during vocal fold abduction that the vocal folds ceased vibration (ie, "last voicing cycle"). The visually-identified last point of contact and the voicing cycles from each voicing offset were then used to calculate RFF via two methods as detailed below. These two methods varied in the criteria used to determine which voicing cycle was considered the "final RFF cycle." Figure 2 shows a schematic of each method.

"Vocal Fold Contact RFF" used the last point of vocal fold contact to determine which cycle of voicing corresponded to the final RFF cycle. In this method, all individual utterances were aligned based on the last point of vocal fold contact, which was a variable distance away from the end of phonation. Specifically, the final RFF cycle was determined as the fifth cycle following the last point of vocal fold contact. The fifth cycle was used, because it was found that, on average, the last point of vocal fold contact and the last point of voicing were five cycles apart. In instances in which there were not five cycles following the last point of vocal fold contact, the utterance was omitted for both RFF methods (see Utterance Rejection Criteria). The final RFF cycle was normalized into ST using a reference cycle, which was calculated as the first full cycle immediately preceding the point in time that corresponded to 0.069s before the last cycle of voicing. This duration corresponded to the average time of ten voicing cycles across all participants in the dataset.

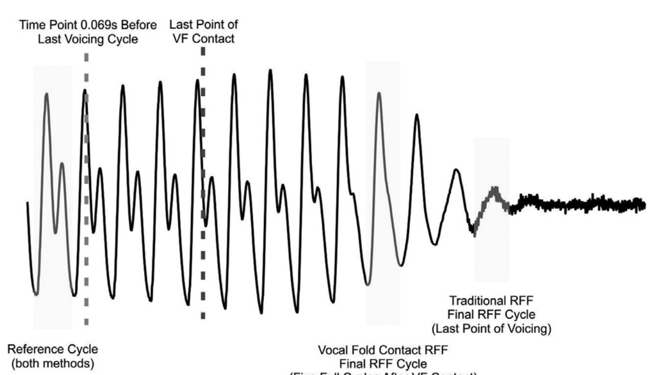

FIGURE 2. Example of identifying the cycles used in Vocal Fold Contact RFF and Traditional RFF to calculate final RFF cycle values in a vowel preceding a voiceless consonant.

"Traditional RFF" was the same as RFF methods found in previous literature [eg,[1]], wherein the final RFF cycle was considered to be the final cycle of voicing (as determined by the automated RFF algorithm in the current study), which was a variable distance away from the end of vocal fold contact. Thus, on average, Vocal Fold Contact RFF and Traditional RFF should use the same cycle as the final RFF cycle, though individual utterances may exhibit variability. Specifically, the final RFF cycle in Vocal Fold Contact RFF may occur earlier in the voicing offset than the final RFF cycle in Traditional RFF, which is always the final cycle of voicing. The final RFF cycle in Traditional RFF was normalized in ST using the same reference cycle as in Vocal Fold Contact RFF to maintain consistency between methods. On average, this should result in 10 total RFF cycles, corresponding to RFF methods from previous literature.

Following RFF calculations via both methods, the within-participant standard deviations (SDs) and means of the final RFF cycle values were calculated for both methods and for both effort conditions. The SDs of the final RFF cycle values were used to evaluate the first research question and provide empirical evidence to support the relationship proposed in.[31] The means of the final RFF cycle values were used to evaluate the second research question and validate that both methods are similarly sensitive to changes in vocal effort.

Utterance rejection criteria

During data analysis, individual utterances were rejected for a number of reasons. First, utterances were rejected if the endoscopic videos were unusable, and the last point of contact could not be properly estimated. This may have occurred if the video was too dark or blurry, or if the speaker's epiglottis covered the vocal folds and a clear image was not obtained. Second, utterances with physiologically invalid RFF values using either method were rejected. Physiologically invalid RFF values were defined using the same parameters as in previously reported automated RFF algorithms.[36,37] Third, utterances with less than five voicing cycles following the last point of contact were rejected. This is because Vocal Fold Contact RFF requires five complete voicing cycles. In order to maintain an identical dataset across methods, these utterances were rejected for both Vocal Fold Contact RFF and Traditional RFF measures such that differences in SDs and means between methods would not be the result of a decrease in the number of usable utterances. Finally, in instances in which utterance rejection resulted in a speaker with less than two usable utterances in either effort condition, the individual was removed prior to statistical analysis, as SDs and means were not able to be calculated for these speakers.

Statistical analysis

All statistical analysis was completed in Minitab with an *a priori* significance level of $P < 0.05$. Two repeated-measures analyses of variance (ANOVAs) were performed with final RFF cycle SD and mean as the outcome variables. “RFF method” and “effort condition” were set as fixed factors and participant was set as a random factor. The interaction effect between RFF method and effort condition was also evaluated. Partial eta-squared (η_p^2) values were calculated and qualitatively interpreted to determine small, medium, or large effect sizes for significant factors.[38]

RESULTS

Utterance rejection

Prior to utterance rejection, a total of 1289 utterances were recorded across all 45 speakers. Of these initial utterances, 270 were rejected due to unusable videos, 34 were rejected due to physiologically invalid RFF values, and 352 were rejected because there were less than five cycles following the last point of contact. This resulted in a final total of 633 utterances. A total of 10 participants were rejected from statistical analysis due to a lack of usable utterances. This resulted in a final N = 35 speakers with an average of 8.5 (standard deviation = 4.0) usable utterances per effort condition. The average RFF patterns across all 35 speakers for both effort conditions using both methods are shown in Figure 3. Only the final RFF cycle was used for statistical analysis.

Within-speaker final RFF cycle SDs and means

The results of the ANOVA investigating within-speaker final RFF cycle SDs are shown in Table 1. Participant ($P < 0.01$) and RFF method ($P < 0.01$) had statistically significant effects on final RFF cycle SDs. Participant had a large effect size ($\eta_p^2 = 0.40$) and RFF method had a medium effect size ($\eta_p^2 = 0.22$), with Vocal Fold Contact RFF resulting in smaller SDs than Traditional RFF. Effort condition and the interaction between RFF method and effort condition did not have statistically significant effects on final RFF cycle SDs. Average within-speaker final RFF cycle SD and 95% confidence intervals as a function of RFF method and effort condition are shown in Figure 4.

The results of the ANOVA investigating within-speaker final RFF cycle means are shown in Table 2. Participant ($P < 0.01$), effort condition ($P < 0.01$), and RFF method ($P = 0.04$) had statistically significant effects on final RFF cycle means. Participant had a large effect size ($\eta_p^2 = 0.72$), effort condition had a medium effect size ($\eta_p^2 = 0.20$; corresponding to a decrease in final RFF cycle means for maximum effort), and RFF method had a small effect size

TABLE 1.
Analysis of Variance Results for Final Relative Fundamental Frequency (RFF) Cycle Standard Deviations, with Participant as a Random Factor, and RFF Method, Effort Condition, and the Interaction Between RFF Method and Effort Condition as Fixed Factors

Factor	*df*	F-value	*P*-value	η_p^2	Effect Size
Participant	34	2.01	< 0.01	0.40	Large
RFF method	1	27.97	< 0.01	0.22	Medium
Effort condition	1	0.29	0.59	–	–
Method x Condition	1	0.11	0.75	–	–

Notes: Dashes indicate non-significant findings.
Abbreviations: df, degrees of freedom; η_p^2, partial eta squared value for evaluation of effect size.

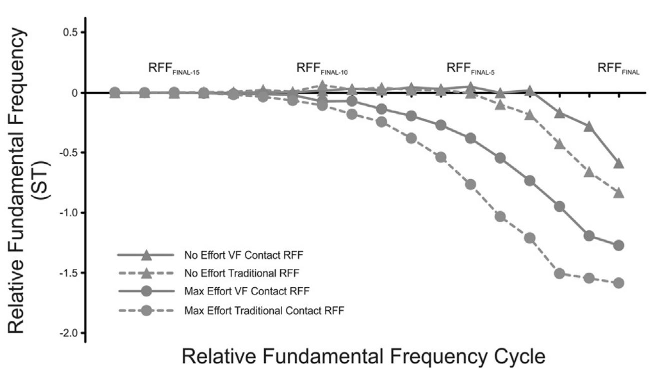

FIGURE 3. Relative fundamental frequency (RFF) patterns (measured in semitones) averaged across all 35 participants for both effort conditions (no effort and maximum effort) using both methods (Vocal Fold Contact RFF and Traditional RFF). Individual utterances were aligned by the Final RFF Cycle (RFF_{FINAL}) and calculated for all preceding cycles up to the reference cycle. VF Contact RFF, vocal fold contact RFF.

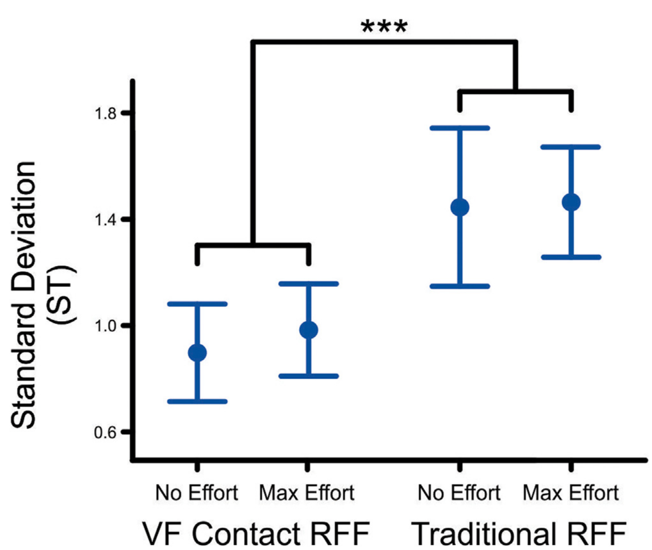

FIGURE 4. Average and 95% confidence intervals for the final cycle relative fundamental frequency (RFF) standard deviation (in semitones) as a function of RFF method and effort condition. *** indicates a statistically significant difference of ($P < 0.05$). VF Contact RFF, vocal fold contact RFF.

TABLE 2.
Analysis of Variance Results for the Final Relative Fundamental Frequency (RFF) Cycle Means, with Participant as a Random Factor, and RFF Method, Effort Condition, and the Interaction Between RFF Method and Effort Condition as Fixed Factors

Factor	*df*	F-value	*P*-value	η_p^2	Effect Size
Participant	34	7.71	< 0.01	0.72	Large
RFF method	1	4.55	0.04	0.04	Small
Effort condition	1	25.74	< 0.01	0.20	Medium
Method x Condition	1	0.03	0.86	–	–

Notes: Dashes indicate non-significant findings.
Abbreviations: df, degrees of freedom; η_p^2, partial eta squared value for evaluation of effect size.

($\eta_p^2 = 0.04$; corresponding to a decrease in final RFF cycle means for Traditional RFF). The interaction between RFF method and effort condition did not have a statistically significant effect on final RFF cycle means. Average within-speaker final RFF cycle means and confidence intervals as a function of RFF method and effort condition are shown in Figure 5.

DISCUSSION

Vocal fold contact RFF decreases variability

When calculating RFF based on the time at which vocal fold contact ended, RFF variability decreased when compared to the traditional method of calculating RFF (Figure 4). Final RFF cycle SDs decreased by an average of 37.9% for the no effort condition and 33.0% for the maximum effort condition. This supports our hypothesis for the first research question, which sought to investigate how aligning RFF offset based on the last point of vocal fold contact would affect within-speaker variability in the final RFF cycle when compared to traditional RFF aligned to the end of voicing. The research question was based on the relationship proposed by,[31] which suggested that the decrease in typical RFF offset patterns may be due, in part, to the decrease in the collision forces of the vocal folds during abduction, causing a decrease in system stiffness. The results of the current study suggest that standardizing the time between the end of vocal fold contact and the cycle at which RFF is measured across individual utterances decreases RFF variability, thereby supporting the relationship proposed in.[31] Additionally, since effort condition did not have a statistically significant effect on within-speaker SDs, it is likely that this relationship holds true regardless of changes in vocal effort.

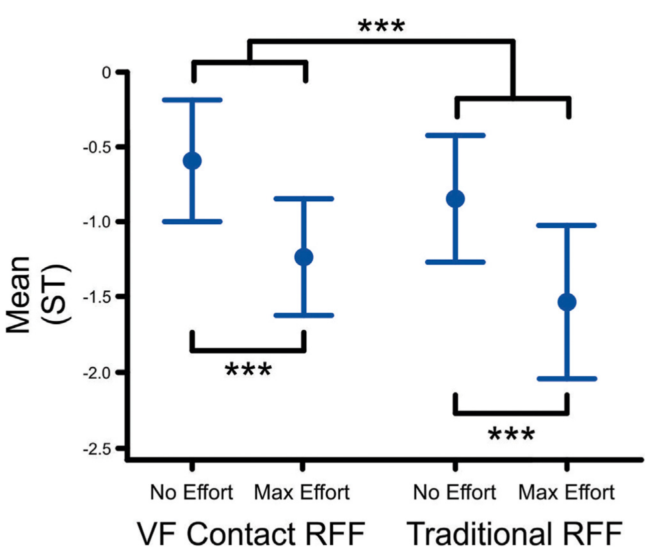

FIGURE 5. Average and 95% confidence intervals for the final cycle relative fundamental frequency (RFF) means (in semitones) as a function of RFF method and effort condition. *** indicates a statistically significant difference of ($P < 0.05$). VF Contact RFF, vocal fold contact RFF.

Vocal fold contact RFF decreases with increased vocal effort

RFF means were significantly lower for the maximum vocal effort condition when compared to the no vocal effort condition. This indicates that RFF means decreased when speakers with typical voices were instructed to increase their vocal effort (Figure 5), which aligns with previous work that showed that RFF measures decrease when estimates of laryngeal tension increase.[2,3] There was also no significant interaction effect between RFF Method and Condition, demonstrating that there was a decrease in RFF means across effort condition for both methods of calculating RFF. This supports our hypothesis for the second research question, which stated that both Vocal Fold Contact RFF and Traditional RFF would show a decrease in RFF means when individuals increased their vocal effort. The results of the current study indicate that the new method of RFF, Vocal Fold Contact RFF, successfully demonstrates sensitivity to changes in vocal effort.

Despite both RFF methods showing sensitivity to changes in vocal effort, RFF method had a small, but statistically significant effect on RFF mean. Vocal Fold Contact RFF resulted in statistically significantly higher RFF means than Traditional RFF. Further, the average change in RFF means between effort conditions when using Vocal Fold Contact RFF (0.64 ST) was somewhat smaller than when using Traditional RFF (0.69 ST). A smaller change in Vocal Fold Contact RFF is expected, given that, in some utterances, the final cycle used for Vocal Fold Contact RFF may have occurred earlier in the offset vowel than the final cycle used for Traditional RFF, which is always the final voicing cycle. These earlier cycles are closer to the steady-state reference cycle and have been shown to be less sensitive to changes in vocal effort.[2] This effect could mask subtle changes in RFF, such as the small decrease in RFF observed when individuals speak with mild levels of effort.[2]

However, RFF Method exhibited only a small effect size ($\eta_p{}^2$ = .04[38];) and the average difference between effort conditions was marginal (0.05 ST). This difference is unlikely to have a meaningful impact on measures made with Vocal Fold Contact RFF. For example, in a study that calculated RFF in individuals with vocal hyperfunction before and after voice therapy, RFF offset cycle 10 changed by an average of 0.50 ST,[13] a difference that is unlikely to be masked by using Vocal Fold Contact RFF. Therefore, in addition to validating the physiological mechanisms behind RFF patterns, Vocal Fold Contact RFF shows potential as an acoustic measure that is sensitive to changes in estimates of laryngeal tension.

Study limitations and future work

Although the decreased variability in RFF offset when using Vocal Fold Contact RFF supports the theory that deceases in RFF offset may be due to a decrease in collision forces at the end of vocal fold contact, there are a few limitations to this study that should be acknowledged. Laryngeal images were captured at 1000 Hz to maximize the level of light in the images. However, this means that, depending on the f_o of the speaker, the exact point at which vocal fold contact ended may have been masked by the limited number of frames per cycle. Thus, the time point that was marked as the end of vocal fold contact for each utterance was an estimate of the exact time point. An increased framerate may improve the accuracy of this estimate, but the end of vocal fold contact was only used to determine which full cycle to mark as the final RFF cycle. Male speakers have an average f_o of 85 to 180 Hz, whereas female speakers have an average f_o of 165 to 255 Hz.[39] Even at the highest average female f_o, a sampling rate of 1000 Hz results in each frame representing approximately one fourth of a single cycle. As a result, increasing the framerate would have very little effect on cycle selection.

Differences in average f_o across speakers may also introduce variability to measures of RFF. Speakers with a lower f_o will have larger cycle lengths, resulting in RFF measures over a longer period of time than within speakers with a higher f_o. As a result, it may be better to investigate RFF within speakers with similar average f_o. However, because both methods used to calculate RFF in the current study are affected by this variability, it is unlikely that differences in the cycle lengths which RFF is calculated have a significant effect on the results of this study.

Another methodological limitation of the study is that calculating Vocal Fold Contact RFF requires five complete cycles following the end of vocal fold contact. Any utterance that did not have five complete cycles following the end of vocal fold contact was excluded from both Vocal Fold Contact RFF and Traditional RFF calculations in order to avoid rejection bias that would favor one method over the other. Five complete cycles were chosen, because, on average, the end of vocal fold contact occurred five cycles before the last voicing cycle, meaning that the average final RFF cycle was the same for Vocal Fold Contact RFF and Traditional RFF methods. However, this methodological decision meant that the 36% of all utterances were rejected. This also meant that 10 of the 45 speakers did not have enough usable utterances to calculate means and SDs and were excluded from statistical analysis. Future work should investigate the impact of removing the five cycle rejection criteria in order to reduce the amount of data that would be rejected. Regardless, despite the loss of data in the current study, subsequent analyses were conducted on a total of 35 speakers, which is larger than previous studies that demonstrated changes in RFF when 12 speakers with typical voices were instructed to speak with increased effort.[2,3] Additionally, even after utterance rejection, each speaker had an average of 8.5 utterances per condition, which is larger than the recommended number of six utterances for stable RFF measures.[20] Given that the utterances used were the same for both methods of RFF calculation, the reduction of data likely did not meaningfully affect the results of the study.

The current study utilized a novel method of calculating RFF via alignment by the end of vocal fold contact, which required the use of endoscopy. Despite a decrease in RFF variability, the reliance on endoscopic inspection of the vocal folds to calculate RFF is not clinically feasible. However, the current study incorporated Vocal Fold Contact RFF specifically to investigate the relationship between the end of vocal fold collision and voicing offset to better understand the physiological mechanisms behind RFF offset patterns in an experimental setting. Clinical feasibility was not considered when developing the method of Vocal Fold Contact RFF. Future work should investigate the potential to estimate the end of vocal fold contact directly from the acoustic signal, such that Vocal Fold Contact RFF may be calculated without the use of endoscopy.

The speakers in this study were individuals with typical voices who were instructed to speak with increased vocal effort, instead of individuals with voice disorders characterized by excessive laryngeal tension. Furthermore, although participants were given identical instructions, they may have used different levels of vocal effort during the maximum effort condition. It is possible that the relationship between the end of vocal fold contact and the decrease in RFF offset is different between individuals with typical voices speaking with increased vocal effort and individuals with voice disorders characterized by excessive laryngeal tension. Thus, the results of this study may not be applicable to all speakers. Future studies should investigate this by designing a similar study that investigates Vocal Fold Contact RFF in individuals with voice disorders characterized by excessive laryngeal tension.

Finally, the vocal fold model of[31] specifically investigated changes to RFF during voicing offset. Prior to the start of this study, a similar model had not been developed to investigate voicing onset, so the current study focused only on voicing offset. As future modeling work explores changes in

RFF during voicing onset, these models should subsequently be validated by a similar study.

CONCLUSION

When aligning utterances based on the end of vocal fold contact during abduction, there is a decrease in within-speaker variability of RFF offset measures. This supports the theory that decreases in RFF during voicing offset are due to the reduction in system stiffness when the vocal folds cease to contact during vibration. The results of this study provide important information about the physiological mechanisms behind RFF and may help improve the feasibility of RFF as an acoustic measure for estimating changes in laryngeal tension.

Acknowledgments

Thanks to Manuel Díaz-Cádiz for creating custom MATLAB scripts used for data analysis and to Dante Cilento, Austin Luong, and Laura Raiff for helping with the data analysis.

REFERENCES

1. Stepp CE, Hillman RE, Heaton JT. The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *J Speech Lang Hear Res*. 2010;53:1220–1226.
2. Lien YA, Michener CM, Eadie TL, Stepp CE. Individual monitoring of vocal effort with relative fundamental frequency: relationships with aerodynamics and listener perception. *J Speech Lang Hear Res*. 2015;58:566–575.
3. McKenna VS, Heller Murray ES, Lien YA, Stepp CE. The relationship between relative fundamental frequency and a kinematic estimate of laryngeal stiffness in healthy adults. *J Speech Lang Hear Res*. 2016;59:1283–1294.
4. Watson BC. Fundamental frequency during phonetically governed devoicing in normal young and aged speakers. *J Acoust Soc Am*. 1998;103:3642–3647.
5. Robb MP, Smith AB. Fundamental frequency onset and offset behavior. *J Speech Lang Hear Res*. 2002;45:446–456.
6. Xu Y, Xu A. Consonantal F0 perturbation in American English involves multiple mechanisms. *J Acoust Soc Am*. 2021;149:2877–2895.
7. Kirby JP, Ladd DR. Effects of obstruent voicing on vowel F0: evidence from "true voicing" languages. *J Acoust Soc Am*. 2016;140:2400–2411.
8. Haggard M, Ambler S, Callow M. Pitch as a voicing cue. *J Acoust Soc Am*. 1970;47(2B):613–617.
9. Hanson HM. Effects of obstruent consonants on fundamental frequency at vowel onset in English. *J Acoust Soc Am*. 2009;125:425–441.
10. Heller Murray ES, Lien YA, Van Stan JH, et al. Relative fundamental frequency distinguishes between phonotraumatic and non-phonotraumatic vocal hyperfunction. *J Speech Lang Hear Res*. 2017;60:1507–1515.
11. Lien YA, Gattuccio CI, Stepp CE. Effects of phonetic context on relative fundamental frequency. *J Speech Lang Hear Res*. 2014;57:1259–1267.
12. Stepp CE. Relative fundamental frequency during vocal onset and offset in older speakers with and without Parkinson's disease. *J Acoust Soc Am*. 2013;133:1637–1643.
13. Stepp CE, Merchant GR, Heaton JT, Hillman RE. Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction. *J Speech Lang Hear Res*. 2011;54:1260–1266.
14. Stepp CE, Sawin DE, Eadie TL. The relationship between perception of vocal effort and relative fundamental frequency during voicing offset and onset. *J Speech Lang Hear Res*. 2012;55:1887–1896.
15. Goberman AM, Blomgren M. Fundamental frequency change during offset and onset of voicing in individuals with Parkinson disease. *J Voice*. 2008;22:178–191.
16. Park Y, Stepp CE. The effects of stress type, vowel identity, baseline f (0), and loudness on the relative fundamental frequency of individuals with healthy voices. *J Voice*. 2019;33:603–610.
17. van Mersbergen M, Beckham BH, Hunter EJ. Do we need a measure of vocal effort? Clinician's report of vocal effort in voice patients. *Perspec ASHA Spec Interest Groups*. 2021;6:69–79.
18. Hillman RE, Stepp CE, Van Stan JH, et al. An updated theoretical framework for vocal hyperfunction. *Am J Speech Lang Pathol*. 2020;29:2254–2260.
19. Roy N, Fetrow RA, Merrill RM, Dromey C. Exploring the clinical utility of relative fundamental frequency as an objective measure of vocal hyperfunction. *J Speech Lang Hear Res*. 2016;59:1002–1017.
20. Eadie TL, Stepp CE. Acoustic correlate of vocal effort in spasmodic dysphonia. *Annals Otol Rhinol Laryngol*. 2013;122:169–176.
21. Hunter EJ, Cantor-Cutiva LC, van Leer E, et al. Toward a consensus description of vocal effort, vocal load, vocal loading, and vocal fatigue. *J Speech Lang Hear Res*. 2020;63:509–532.
22. McKenna VS, Stepp CE. The relationship between acoustical and perceptual measures of vocal effort. *J Acoust Soc Am*. 2018;144:1643.
23. Lofqvist A, Baer T, McGarr NS, Story RS. The cricothyroid muscle in voicing control. *J Acoust Soc Am*. 1989;85:1314–1321.
24. Stevens KN. Physics of laryngeal behavior and larynx modes. *Phonetica*. 1977;34:264–279.
25. Van Den Berg J. Myoelastic-aerodynamic theory of voice production. *J Speech Lang Hear Res*. 1958;1:227–244.
26. Ohde RN. Fundamental frequency as an acoustic correlate of stop consonant voicing. *J Acoust Soc Am*. 1984;75:224–230.
27. Löfqvist A, Koenig LL, McGowan RS. Vocal tract aerodynamics in /aCa/utterances: measurements. *Speech Commun*. 1995;16:49–66.
28. Hombert J-M, Ohala JJ, Ewan WG. Phonetic explanations for the development of tones. *Language*. 1979;55:37–58.
29. Ladefoged P. *Elements of Acoustic Phonetics*. 2nd ed. Chicago: University of Chicago Press; 1996.
30. Smith AB, Robb MP. Factors underlying short-term fundamental frequency variation during vocal onset and offset. *Speech Lang Hear*. 2013;16:208–214.
31. Serry MA, Stepp CE, Peterson SD. Physics of phonation offset: towards understanding relative fundamental frequency observations. *J Acoust Soc Am*. 2021;149:3654–3664.
32. Story BH, Titze IR. Voice simulation with a body-cover model of the vocal folds. *J Acoust Soc Am*. 1995;97:1249–1260.
33. Park Y, Wang F, Diaz Cadiz ME, et al. Vocal fold kinematics and relative fundamental frequency as a function of obstruent type and speaker age. *J Acoust Soc Am*. 2021;149:2189–2199.
34. Diaz Cadiz ME, McKenna VS, Vojtech JM, Stepp CE. Adductory vocal fold kinematic trajectories during conventional versus high-speed videoendoscopy. *J Speech Lang Hear Res*. 2019;62:1685–1706.
35. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–163.
36. Vojtech JM, Segina RK, Buckley DP, et al. Refining algorithmic estimation of relative fundamental frequency: accounting for sample characteristics and fundamental frequency estimation method. *J Acoust Soc Am*. 2019;146:3184.
37. Lien YA, Heller Murray ES, Calabrese CR, et al. Validation of an algorithm for semi-automated estimation of voice relative fundamental frequency. *Ann Otol Rhinol Laryngol*. 2017;126:712–716.
38. Witte RS, Witte JS. *Statistics*. Hoboken, NJ: J. Wiley & Sons; 2010.
39. Titze IR. *Principles of Voice Production*. Englewood Cliffs, N.J.: Prentice Hall; 1994.