JSLHR

Research Article

The Relationship Between Voice Onset Time and Increase in Vocal Effort and Fundamental Frequency

Matti D. Groll,[a,b] Surbhi Hablani,[b] and Cara E. Stepp[a,b,c]

Purpose: Prior work suggests that voice onset time (VOT) may be impacted by laryngeal tension: VOT means decrease when individuals with typical voices increase their fundamental frequency (f_o) and VOT variability is increased in individuals with vocal hyperfunction, a voice disorder characterized by increased laryngeal tension. This study further explored the relationship between VOT and laryngeal tension during increased f_o, vocal effort, and vocal strain.
Method: Sixteen typical speakers of American English were instructed to produce VOT utterances under four conditions: baseline, high pitch, effort, and strain. Repeated-measures analysis of variance models were used to analyze the effects of condition on VOT means and standard deviations (*SD*s); pairwise comparisons were used to determine significant differences between conditions.
Results: Voicing, condition, and their interaction significantly affected VOT means. Voiceless VOT means significantly decreased for high pitch ($p < .001$) relative to baseline; however, no changes in voiceless VOT means were found for effort or strain relative to baseline. Although condition had a significant effect on VOT *SD*s, there were no significant differences between effort, strain, and high pitch conditions relative to baseline.
Conclusions: Speakers with typical voices likely engage different musculature to increase pitch than to increase vocal effort and strain. The increased VOT variability present with vocal hyperfunction is not seen in individuals with typical voices using increased effort and strain, supporting the assertion that this feature of vocal hyperfunction may be related to disordered vocal motor control rather than resulting from effortful voice production.

Voice onset time (VOT) is an acoustic measure of the interval of time between the release of the oral constriction in a stop consonant and the onset of subsequent voicing (Lisker & Abramson, 1964). In American English, VOT is used by listeners to distinguish whether a consonant is perceived as "voiced" (Zlatin, 1974). Voiced stop consonants (/bdg/) are produced with a short lag, and voiceless stop consonants (/ptk/) are produced with a long lag. Although it is possible for voicing to occur prior to the plosive release, resulting in a negative VOT, this is primarily unused in the English language (Klatt, 1975; Rae, 2018).

VOT is affected by a variety of factors, including speaking rate (Allen et al., 2003; Kessinger & Blumstein, 1997, 1998; Volaitis & Miller, 1992), age (Sweeting & Baken, 1982; Whiteside et al., 2003; Yu et al., 2015; Zlatin & Koenigsknecht, 1976), sex (Swartz, 1992), phonetic context (Klatt, 1975; Morris et al., 2008), native language of the speaker (Lisker & Abramson, 1964; Narayan & Bowden, 2013), lung volume (Hoit et al., 1993), and the presence of communication disorders (Ackermann & Hertrich, 1997; Whitfield & Goberman, 2015). Previous work has used a variety of acoustic stimuli and methodologies, which makes it difficult to specify normalized values. However, in a review of VOT literature, Auzou et al. compiled the mean VOT values for American English stops from a number of studies. Mean VOT values reported for voiceless stops ranged from roughly 40 to 110 ms and 0–25 ms for voiced stops (Auzou et al., 2000). Due to the smaller range of acceptable durations for voiced stops, it is common to only observe VOT differences as a result of voice characteristics (e.g., speaking rate) in voiceless stops (Allen et al., 2003;

[a]Department of Biomedical Engineering, Boston University, MA
[b]Department of Speech, Language, and Hearing Sciences, Boston University, MA
[c]Department of Otolaryngology – Head and Neck Surgery, Boston University School of Medicine, MA
Correspondence to Matti D. Groll: mgroll@bu.edu

Editor-in-Chief: Bharath Chandrasekaran
Editor: Jack J. Jiang
Received August 26, 2020
Revision received October 19, 2020
Accepted January 13, 2021
https://doi.org/10.1044/2021_JSLHR-20-00505

***Disclosure**: Cara E. Stepp has received consulting fees from Altec, Inc./Delsys, Inc., companies focused on developing and commercializing technologies related to human movement. Stepp's interests were reviewed and are managed by Boston University in accordance with their conflict of interest policies. All other authors have declared that no other competing interests existed at the time of publication.*

Kessinger & Blumstein, 1997; McCrea & Morris, 2005; Narayan & Bowden, 2013).

Whereas the effects of some factors on VOT are well studied (e.g., age; Sweeting & Baken, 1982; Whiteside et al., 2003; Yu et al., 2015; Zlatin & Koenigsknecht, 1976), research on the effects of other factors remains limited. One such factor is fundamental frequency (f_o). During voicing, vocal fold vibrations result from changes to air pressure and flow in the vocal tract (Titze, 1988). It has been well established that elevating the larynx is associated with increased pitch (Ohala, 1977). This gesture decreases the intraoral volume, causing a reduction in the pressure drop across the glottis, thereby reducing the airflow through the glottis (Stevens, 1977). A reduction of airflow is thought to inhibit vocal fold vibration and extend the initiation of voicing, resulting in a longer VOT (Stevens, 1977). Furthermore, the necessary subglottal pressure required to initiate vocal fold vibration, known as the phonation threshold pressure, has been shown to increase as f_o increases (Titze, 1992). Thus, at increased f_o, it follows that the time needed to achieve a higher phonation threshold pressure will increase, resulting in a longer VOT. In summary, aerodynamic theories of voice production predict that increases in vocal f_o should lead to longer VOTs.

Contrary to the theory that increased f_o would result in longer VOTs, McCrea and Morris (2005) found that the VOTs of voiceless stop consonants were significantly shorter when 60 adult men with typical voices spoke at a higher f_o. The authors reasoned that the shorter VOTs found at the higher f_o values were a result of increased levels of vocal fold stiffness that affected vibrational characteristics as well as abductory and adductory gestures. A speaker's f_o is dependent on the rate of vocal fold vibration and is therefore affected by the length, mass, and tension of the vocal folds (Van Den Berg, 1958). It is believed that increases in f_o are achieved by stiffening the vocal folds through the recruitment and increased contraction of the cricothyroid, thyroarytenoid, and suprahyoid muscles (Stemple et al., 2018). A higher degree of vocal fold stiffness also allows the vocal folds to quickly return from the abducted position during a stop consonant burst to the adducted position of subsequent voicing, which would result in a shorter VOT. More recent work showed a similar decrease in VOT when adult women with typical voices increased f_o (Narayan & Bowden, 2013), adding potential support to the idea that a shorter VOT may be a result of increased vocal fold stiffness. If this argument is true, then increased tension in the intrinsic laryngeal muscles responsible for vocal fold stiffening should also result in a shorter VOT irrespective of pitch increases. Weismer (1984) observed that individuals with Parkinson's disease, often described as having increased tension (rigidity) of the laryngeal musculature, have shorter voiceless VOT values when compared to older and younger speakers with typical voices (Weismer, 1984), lending further credence to this hypothesis. However, Parkinson's disease has various complex effects on voice production; the shorter voiceless VOT values in individuals with Parkinson's disease are not necessarily attributed to increased laryngeal tension.

To further explore the relationship between laryngeal tension and VOT, McKenna et al. (2020) investigated VOT in individuals with vocal hyperfunction, defined classically as impaired "vocal mechanisms as a result of excessive or imbalanced muscle activity" (Hillman et al., 1989). Vocal hyperfunction is a primary characteristic of muscle tension dysphonia as well as organic voice disorders such as polyps and vocal fold nodules (Mehta et al., 2015). McKenna et al. (2020) compared 32 females with vocal hyperfunction to 32 age- and sex-matched controls. The authors found that there was no difference in mean VOT between groups (McKenna et al., 2020). These results could suggest that the f_o-based difference in VOT found by McCrea and Morris (2005) is due to a mechanism other than an increase in vocal fold stiffness. However, these results may also be attributed to a lack of increased vocal fold stiffness in this sample of individuals with vocal hyperfunction, because vocal hyperfunction does not necessarily specify tension in the muscles responsible for vocal fold stiffening, but instead is characterized as general intrinsic and extrinsic muscular tension (Aronson, 1990). Thus, these participants may have exhibited overall increased laryngeal tension despite not presenting excess tension in the muscles specific to vocal fold stiffening (i.e., the cricothyroid and the thyroarytenoid muscles). Alternatively, the laryngeal tension in this group of individuals with vocal hyperfunction may simply not have been elevated enough to significantly affect VOT. The authors proposed that these individuals only exhibited mild hyperfunction and that individuals with greater levels of hyperfunction may have had an effect on VOT (McKenna et al., 2020). This explanation could be supported by the results of McCrea and Morris, who found a significant difference in VOT between low and high pitch levels, but not between low and medium pitch levels (McCrea & Morris, 2005). Regardless of the interpretation, more work with individuals who explicitly have laryngeal tension is necessary to fully understand the relationship between laryngeal tension and VOT measures.

McKenna et al. (2020) also observed that, within an individual, the VOTs of speakers with vocal hyperfunction were significantly more variable than in speakers with typical voices for both voiced and voiceless stops (McKenna et al., 2020). The authors reasoned that this could be attributed to an impairment in their speech motor control. Contemporary models of speech motor control, including the Directions Into Velocities of Articulators model (Tourville & Guenther, 2011), theorize that, during development of speech, the auditory control system relies primarily on a feedback control system to monitor and correct motor behaviors in order to produce speech within a specific target area. In the developed adult system, after speech target areas have been learned, the speech motor control system relies primarily on feedforward control to produce speech within the desired target area, though auditory feedback is still used to monitor and correct errors. A study by Stepp et al. (2017) suggested that a subset of individuals with vocal hyperfunction may have a predisposition to auditory-motor integration deficits, which results in a maladaptive updating of the feedforward system (Stepp et al., 2017). In other words, individuals with vocal hyperfunction tend to improperly update their feedforward system in response to perturbations in auditory feedback. This may result in a larger target area during speech production. Since VOT is used to distinguish voicing in English phonemes (Zlatin, 1974) and, therefore, specifies a desired speech target, a larger target area would likely result in a larger range of acceptable VOT values, thereby increasing VOT variability, as observed by McKenna et al. (2020). Based on this interpretation, increased variability would be expected to occur in individuals with vocal hyperfunction irrespective of the momentary degree of laryngeal tension. Furthermore, this increased variability would not be expected in speakers without vocal hyperfunction, even when using increased laryngeal tension.

A limited number of studies have investigated within-subject VOT variability and shown that it is increased in prelingually deaf speakers (Lane & Perkell, 2005), as well as in young children (Whiteside et al., 2003; Yu et al., 2015). However, studies investigating VOT within typical speakers (e.g., McCrea & Morris, 2005; Narayan & Bowden, 2013) do not explicitly report within-subject variability. A lack of reporting of within-subject variability in typical speakers makes it difficult to discern whether the increased VOT variability as observed by McKenna et al. is specific to a potential sensorimotor disorder in individuals with vocal hyperfunction. For example, a general increase in tension in the laryngeal system could be seen as a change from normal laryngeal function, which might result in variability during adductory gestures and, subsequently, more variable VOT measures. If so, VOT variability would increase in all cases of increased laryngeal tension, not just in individuals with vocal hyperfunction. However, in the upper limb, increased muscular stiffness has been shown to result in decreased variability (Lametti & Ostry, 2010). Though it is unclear how these results translate to the laryngeal system, it is likely that an increase in laryngeal tension alone (i.e., without the proposed sensorimotor impairment observed in individuals with vocal hyperfunction) would not cause an increase in VOT variability. Thus, there is currently a need to further investigate VOT variability in individuals expressing increased laryngeal tension.

Increased laryngeal tension is difficult to measure directly and is often associated in literature with both vocal effort and vocal strain (McKenna et al., 2019; McKenna & Stepp, 2018; Rosenthal et al., 2014). In a recent literature review, a formal definition of vocal effort was defined as "the perceived exertion of a vocalist's response to a perceived communication scenario by the speaker" (Hunter et al., 2020). This description of vocal effort indicates, by definition, that it is a feature of the speaker's experience. In contrast, vocal strain is traditionally defined as the auditory-perception of vocal effort by a listener (Kempster et al., 2009). Though these terms are often used concurrently, vocal strain may encompass additional voice qualities such as breathiness and roughness (Lowell et al., 2012). Furthermore, self- and listener perception of vocal effort has demonstrated a moderate relationship (Eadie et al., 2010), with speakers often reporting higher levels of vocal effort than those perceived by listeners (Lane et al., 1961). This difference is to be expected, given that speakers have additional cues related to their perceived effort (such as somatosensory inputs), which are not reflected to listeners by the acoustic signal (Eadie et al., 2010). Thus, although both of these features are thought to be related to increased laryngeal tension, vocal effort and vocal strain may provide differential information about laryngeal function, indicating that both features should be assessed to better investigate increases in laryngeal tension.

Research Statement

We sought to determine the effects of increased f_o, increased vocal effort, and increased vocal strain on VOT measures. Our goal was to better understand the relationships between laryngeal tension and VOT measures suggested by previous studies. Mean VOT and VOT variability, as measured by the standard deviation (*SD*), were calculated for male and female speakers with typical voices who were asked to speak with increased f_o, increased vocal effort, and increased vocal strain. We hypothesized that increased f_o, increased vocal effort, and increased vocal strain would lead to shorter VOT means for voiceless stop consonants. In addition, based on the interpretation that increased variance is a result of a sensorimotor disorder attributed to vocal hyperfunction, we predicted that VOT *SD*s would not change with increased f_o, increased effort, or increased vocal strain. We tested the alternative hypothesis that increased f_o, increased vocal effort, and increased vocal strain would result in an increase in VOT *SD*s.

Method

Participants

A total of $N = 16$ cisgender speakers (eight women, eight men; $M = 21.8$ years, $SD = 2.9$ years) participated in this study. All were native speakers of American English, were nonsmokers, and reported no history of neurological, speech, language, hearing, or voice disorders. Participants gave written informed consent, in compliance with the Boston University Institutional Review Board, and completed a hearing screening prior to the start of the study. A series of three pulsed tones were presented at a range of frequencies from 125 to 8000 Hz with the instructions to indicate when a tone was detected. Tones were presented monaurally at 25 dB HL through an overear headset, first in the right ear and then the left. If a participant failed to identify the tones at a given frequency, the presentation level was increased by 5 dB until the participant identified the tone. All participants identified pulsed tones at all frequencies at 25 dB HL level, except one participant who required a 30 dB HL level for one ear at 4000 Hz.

Experimental Design

Following consent and a hearing screening, participants were instructed to repeat a set of VOT stimuli using

/a/ and /u/ vowels under four different conditions: baseline, high pitch, effort, and strain. Prior to each condition, participants were trained on how they should produce the VOT stimuli. During the baseline condition, participants were instructed to repeat VOT stimuli with a natural speaking voice.

For the high pitch condition, participants were instructed to repeat VOT stimuli with a f_o that was eight semitones higher than their baseline average. In preliminary testing, a target of eight semitones higher than the baseline average was determined to be the greatest increase in semitones that all participants could reach without shifting from the modal register. The average f_o from the baseline condition was calculated (in Hertz) using the Praat acoustic analysis software (Version 6.0.46) for each participant (Boersma, 2001). Using a speech synthesizer program (Madde, Tolvan Data, Version 3.0.0.2), a sustained vowel was played for the participant at the f_o that the participant was instructed to match. Participants were given the opportunity to practice matching the pitch during a training session prior to recording by first matching with the vowel /a/ and then with a subset of sentences from the stimuli set. The experimenter listened and corrected for any errors in matching the appropriate pitch. All participants remained within the modal register at the increased f_o. The vowel to match was repeated immediately before each recording.

For the effort condition, participants were instructed to speak with maximum vocal effort. Experiential anchoring has been shown to improve the validity of effort scales (Lamb et al., 2008). This anchoring can be performed by instructing participants to complete a task at the desired effort level or by recalling a memory of using the desired effort level. In a previous study that examined the use of experiential anchoring on vocal effort ratings, individuals with vocal hyperfunction were told to recall attempting to talk during severe laryngitis as an anchor for maximum effort (Van Leer & Van Mersbergen, 2017). It was reported that all participants verbalized that they had experienced this level of effort in their life. However, preliminary testing in the current study revealed that many individuals without voice disorders had not experienced maximum effort as a result of laryngitis. As a result, participants in the current study were given the following verbal description of vocal effort, adapted from van Leer and van Mersbergen (2017): "Now I would like you to speak with as much vocal effort as you can, while still maintaining a voice. Vocal effort is defined as the perceived amount of effort that it takes to produce your voice, but it is independent of your volume. Vocal effort can be associated with multiple areas of the body that contribute to speech production. Think about a time that you had laryngitis or you were very sick, or after you have a long day of continuous speaking such as presenting in class or shouting at a sporting event. It may have been difficult to produce a voice so you had to use more effort to push your voice out. This is what would be considered maximum effort." In order to ensure that the participant was self-evaluating their vocal effort using experiential anchoring instead of "matching their voice" to the researcher's, no auditory example was provided to the participants. Participants practiced speaking with effort prior to recording. During this training, participants were instructed to not increase their volume or pitch during increased vocal effort.

For the strain condition, participants were instructed to speak with increased vocal strain. Participants were presented with two examples (one male and one female) of individuals speaking with increased strain and given the following verbal instructions: "Strain is the auditory perception of effort. It is the quality of voice that you listen to in order to determine whether someone is using vocal effort. Different individuals will produce a strained voice in different ways. I will play two recordings of someone speaking with strain. Listen carefully to the recordings and then I will ask you to match that voice quality. Note that you are attempting to match only the strained voice quality of the speakers. Your pitch and loudness should remain the same." Strain samples were presented binaurally through a set of headphones (Sennheiser HD280 Pro). Participants were allowed to relisten to each sample upon request. Each sample was produced by an individual with a typical voice imitating the presence of vocal strain by producing the same VOT stimuli used in the current study. The presence of moderate-to-severe strain in each sample was validated by two voice-specializing speech-language pathologists (SLPs). Participants practiced speaking with strain prior to recording.

Across all four conditions, participants were presented with the same stimulus set, consisting of sentences in the format of "Say /vowel-consonant-vowel/ again" for every English stop consonant. The vowels surrounding each stop consonant were either /a/ or /u/. Each unique sentence was repeated 3 times, resulting in a set of 36 sentences (6 consonants × 2 vowels × 3 repetitions). The set of stimuli were separated into four blocks: voiceless stops with vowel /a/, voiced stops with vowel /a/, voiceless stops with vowel /u/, and voiced stops with vowel /u/. Stimuli blocks with the same vowel were always presented sequentially, but the presentation order of vowels and of voicing within each vowel was counterbalanced across conditions within subject. In order to compare the differences between effort and strain and to prevent the strain samples from influencing the effort condition, participants always completed the four conditions in the same order.

Following each condition, participants were asked to rate the amount of vocal effort used to produce the speech stimuli on a 100-mm visual analog scale with anchors of *No Effort* and *Maximum Effort* to the left and right, respectively. A visual analog scale allows for explicit anchors (Gerratt et al., 1993) and can be used to compare self-ratings of vocal effort to listener-perceptual ratings of vocal strain (McKenna & Stepp, 2018). The concept of vocal effort and the maximum effort anchor were explained using the same prompt as during the instructions for the effort condition.

The set of speech stimuli was presented with a custom interface in MATLAB R2018a (MathWorks). A scrolling window was used to highlight sentences at a constant rate in order to maintain a constant speaking rate. Subjects were instructed to read sentences as they were highlighted on the screen. The scrolling rate was set to match a rate of roughly three syllables per second to approximate a typical speaking rate (McCrea & Morris, 2005; Narayan & Bowden, 2013). Each set of speech stimuli took about 90 s to complete. The entire study, including consent, hearing screening, and training, was completed by all participants in under 45 min.

Instrumentation and Calibration

A directional headset microphone (Shure SM35 XLR) was placed 45° from the midline and 7 cm from the lips. A neck-surface accelerometer (BU series 21771 from Knowles Electronic) was placed on the anterior neck, superior to the thyroid notch and inferior to the cricoid cartilage using double-sided adhesive tape and held in place with medical tape. Microphone and Accelerometer signals were pre-amplified (RME Quadmic II) and sampled at 44100 Hz with 16-bit resolution (MOTU UltraLite-mk3 Hybrid).

In order to determine the sound pressure level (SPL) of each sample during data analysis, SPL calibration was completed prior to the baseline condition. An electrolarynx was held in front of the participant's mouth, and an SPL meter was placed next to the headset microphone. The dB SPL values of three levels of the electrolarynx output were recorded in order to determine a relationship between SPL and the amplitude of the recorded acoustic signal.

Data Analysis

VOT was measured manually by a single trained technician for each vowel–consonant–vowel (/vcv/) utterance using a custom MATLAB graphical user interface. The VOT for each utterance was selected by locating the burst of the stop consonant in the microphone signal and the first cycle of voicing following the burst in the accelerometer signal. The accelerometer signal was used to identify the start of voicing, because microphone signals are susceptible to environmental noise and high-frequency aspiration noise as a result of plosives, whereas accelerometer signals measure vibrations through the surface of the neck and are less likely to be impacted by additional noise (Hillman et al., 2006). In the absence of a clear burst in the waveform of the microphone signal, a spectrogram was used to identify the location of the burst. The time period between the burst and the first cycle of voicing was defined as the VOT for that utterance. If the burst of the stop consonant could not be identified or there was no period of devoicing, the utterance was rejected from analysis. VOTs were averaged across usable repetitions to get a mean VOT for each unique /vcv/ utterance (2 vowels × 6 consonants = 12 unique utterances). The VOTs were also used to calculate *SD* for each unique /vcv/ utterance.

Experimental Fidelity

Post hoc analyses were performed to assess the fidelity of the experimental procedures. Participants were instructed to maintain a constant SPL throughout each condition, because SPL has been suggested to have an effect on VOT measures (Knuttila, 2011). In order to verify a constant SPL, the average SPL for each condition was calculated. Similarly, speaking rate has been shown to have a significant effect on VOT measures (Allen et al., 2003; Kessinger & Blumstein, 1997, 1998; Volaitis & Miller, 1992). Although stimuli sentences were presented at a constant rate in order to encourage participants to use a constant speaking rate, it is possible that, in some experimental conditions, participants may have sped up their speaking rate and then paused until the next stimuli sentence was presented. In order to confirm consistent speaking rates across conditions, the average syllable length was calculated for each condition. Audacity software was used to remove long pauses, and syllable length was calculated by dividing the total number of syllables by the length of each audio sample.

Further analysis was used to confirm that participants changed their voice in each experimental condition. In order to verify that participants increased their f_o, average f_o values for the baseline and high pitch conditions were calculated using Praat. Increases in f_o from baseline to the high pitch condition were converted to semitones to confirm an average increase across all participants. Additionally, f_o was calculated for the effort and strain conditions to determine if there were noticeable changes from baseline. To evaluate changes in vocal effort, participants recorded their vocal effort on a visual analog scale from 0 to 100 mm following each condition. The difference between vocal effort ratings from the baseline and effort conditions was used to verify an increase in vocal effort. To assess whether participants increased their vocal strain, two voice-specializing SLPs, blinded to participant and condition, listened to each participant's recordings during each experimental condition and rated strain on a 100-mm visual analog scale according to the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster et al., 2009). Strain ratings were evaluated per speaker using the entire recording for each condition. Vocal strain was reported as an average of the two ratings. The difference between vocal strain ratings from the baseline and strain conditions was used to verify an increase in vocal strain.

Statistical Analysis

In order to estimate interrater reliability for VOT measurements, an independent trained technician remeasured VOTs for three randomly selected participants (18.8% of all data). A two-way intraclass correlation (ICC) analysis was calculated between the two raters. Similarly, three randomly selected participants were reanalyzed by the original trained technician to determine intrarater reliability. During reanalysis, the rater was kept blind to participant and condition. Individual VOT measures for each utterance were compared to the rater's initial measures to determine reliability. The ICC values for interrater and intrarater reliability were ICC(2, 1) = .98 and .99, respectively, indicating excellent reliability for individual VOT measurements (Koo & Li, 2016).

Interrater and intrarater reliability were also calculated for CAPE-V strain ratings. During initial perceptual assessment, 25% of the data were presented to each rater twice. Raters were blinded to participant and condition, as well as whether the sample had been heard before. The ICC value for interrater reliability was ICC(2, 1) = .72, indicating good reliability, and intrarater reliability for each rater was ICC(2, 1) = .93 and .87, indicating excellent and good reliability, respectively (Koo & Li, 2016).

All statistical analyses were completed using Minitab Statistical Software (Version 17; Minitab Inc.). To address the first hypothesis, an analysis of variance (ANOVA) on VOT means was performed. The main effects were condition (baseline, high pitch, effort, and strain), voicing (voiced, voiceless), vowel (/a/, /u/), and place of articulation (bilabial, alveolar, velar). Voicing, vowel, place of articulation, and the two-way interactions between these variables and condition were used to evaluate and confirm the effects of these variables on VOT means that have been reported in previous studies (McCrea & Morris, 2005; McKenna et al., 2020; Morris et al., 2008; Narayan & Bowden, 2013). To address the second hypothesis, a second ANOVA on VOT *SD*s was performed with the same main effects and interactions. Statistical testing was set a priori at $p < .05$, and partial eta squared (η_p^2) was used to determine effect size for each significant effect, with η_p^2 values of .01, .09, and .25 corresponding to small, medium, and large effect sizes, respectively (Witte & Witte, 2010). Tukey's post hoc pairwise comparisons were performed as appropriate. Cohen's *d* was used to determine effect sizes of significant differences, with Cohen's *d* values of 0.2, 0.5, and 0.8 corresponding to small, medium, and large effect sizes, respectively (Witte & Witte, 2010).

Results

VOT means and *SD*s are shown in Table 1. Of the 2,304 VOT utterances, 142 were rejected during analysis, resulting in an average 2.82 utterances per VOT mean and *SD* measure. In instances in which two of the three utterances were rejected, *SD* data were missing (19 instances, 2.5% of data). In instances in which all three utterances were rejected, both VOT mean and *SD* data were missing (12 instances, 1.6% of data).

The results of the two ANOVAs are shown in Table 2. For VOT means, there were significant main effects of condition with small effect size (η_p^2 = .05), place of articulation and vowel with medium effect sizes (η_p^2 = .16, .15, respectively), and voicing with a large effect size (η_p^2 = .81). Additionally, participant had a significant effect on VOT means with a large effect size (η_p^2 = .30). The interaction effect of condition and voicing was significant with a small effect size (η_p^2 = .03). Due to the interaction effect of condition and voicing, Tukey's post hoc pairwise comparison tests were performed to determine significant differences in VOT means between conditions for voiced and voiceless VOT utterances. For voiceless VOT means, values for the high pitch condition were significantly shorter than those in the baseline, effort, and strain conditions, with Cohen's *d* values of 0.48, 0.53, and 0.46, respectively, indicating medium effect sizes (Witte & Witte, 2010). No other conditions were significantly different from one another. Interval plots of voiceless VOT means per condition are shown in Figure 1. For voiced VOT means, there were no significant differences across conditions.

There were significant main effects on VOT *SD*s of condition and place of articulation with small effect sizes (η_p^2 = .02, .01, respectively) and voicing with a medium effect size (η_p^2 = .20). Participant also has a significant effect with a small effect size (η_p^2 = .03). No interaction effects were significant. Tukey's post hoc pairwise comparison tests were used to determine significant differences in VOT *SD*s between conditions for all VOT utterances. VOT *SD*s were significantly lower in high pitch than in effort and strain conditions, with Cohen's *d* values of 0.29 and 0.34, respectively, indicating small effect sizes (Witte & Witte, 2010). However, VOT *SD*s in high pitch, effort, and strain conditions were not significantly different from baseline. Interval plots of VOT *SD*s per condition are shown in Figure 2.

Average SPL and speaking rate across condition were used to investigate experimental fidelity. Means and *SD* for average SPL and speaking rate across condition are reported in Table 3. Changes in f_o, effort ratings, and strain ratings were used to confirm that participants appropriately changed their voice in each experimental condition. The means and *SD* for each variable across all experimental conditions are shown in Table 3.

Discussion

VOT measures in the current study, as shown in Table 1, are consistent with previous literature. Although the method of measuring VOT can differ across studies (for a review, see Rae, 2018), all VOT means were within the range of previously reported norms (Auzou et al., 2000). Compared to (McKenna et al., 2020), in which the same method was used, VOT means were comparable in the current study (23 ms for voiced and 79 ms for voiceless compared to 18 ms for voiced and 64 ms for voiceless). In order to compare VOT variances between the current study and the McKenna study, coefficients of variance, defined as the *SD* divided by the mean, were calculated. These variance values were found to be consistent with values observed in the McKenna study. Furthermore, calculating coefficients of variance and performing identical statistical analysis did not change the primary results of the study.

Changes in VOT Means Across Experimental Conditions

The decrease in VOT means in voiceless productions as a result of increased f_o, as shown in Figure 1, supported our first hypothesis and was consistent with previous studies (McCrea & Morris, 2005; Narayan & Bowden, 2013). Decreases in VOT means while speaking with increased f_o have been suggested (Titze et al., 1988) to be a result of a stiffening of the vocal folds via the activation of the

Table 1. Mean (and standard deviation) of voice onset time (VOT) means and standard deviations (*SD*s) for each condition (baseline, high pitch, effort, strain), as well for all conditions combined (All).

VOT utterance type		VOT means (ms) All	Baseline	High pitch	Effort	Strain
Voicing	Voiced	23 (10)	24 (10)	22 (8)	22 (10)	24 (11)
	Voiceless	79 (24)	81 (25)	70 (22)	83 (23)	83 (26)
Vowel	/a/	46 (32)	47 (33)	41 (27)	48 (35)	47 (34)
	/u/	56 (34)	59 (35)	52 (31)	57 (35)	60 (37)
Place of articulation	Bilabial	43 (32)	43 (33)	41 (27)	45 (33)	45 (35)
	Alveolar	55 (36)	56 (36)	49 (32)	56 (40)	58 (38)
	Velar	56 (32)	59 (33)	49 (28)	56 (32)	58 (33)

VOT utterance type		VOT *SD*s (ms) All	Baseline	High pitch	Effort	Strain
Voicing	Voiced	3.7 (3.9)	4.0 (5.5)	2.8 (2.1)	3.9 (3.5)	4.1 (3.3)
	Voiceless	8.6 (6.2)	8.0 (5.8)	7.4 (4.9)	9.3 (7.2)	9.9 (6.5)
Vowel	/a/	5.7 (6.1)	5.2 (5.5)	4.8 (4.7)	6.5 (7.2)	6.4 (6.3)
	/u/	6.7 (5.4)	6.8 (6.3)	5.6 (4.2)	6.7 (5.2)	7.7 (5.5)
Place of articulation	Bilabial	6.4 (6.1)	6.5 (6.6)	5.4 (5.2)	6.5 (6.1)	7.1 (6.2)
	Alveolar	6.2 (6.1)	5.4 (5.6)	4.9 (4.2)	6.7 (6.9)	7.6 (7.0)
	Velar	6.2 (5.0)	6.2 (5.7)	5.4 (4.0)	6.6 (5.7)	6.5 (4.3)

cricothyroid and thyroarytenoid muscles (Stemple et al., 2018). This increase in vocal fold stiffness decreases the glottal width and could therefore reduce the amount of time needed to return from the abducted position during the voiceless plosive back to the adducted position during the subsequent voicing, thus decreasing VOT means (McCrea & Morris, 2005). As expected, participants had a large effect on VOT means: Average VOT values have been shown to vary significantly across different speakers (Auzou et al., 2000). It is also unsurprising that a difference in VOT means was only observed in voiceless stops. Voiced stops (/bdg/) have a smaller range of acceptable VOTs and are therefore less likely to be impacted by changes in laryngeal tension (McCrea & Morris, 2005; McKenna et al., 2020).

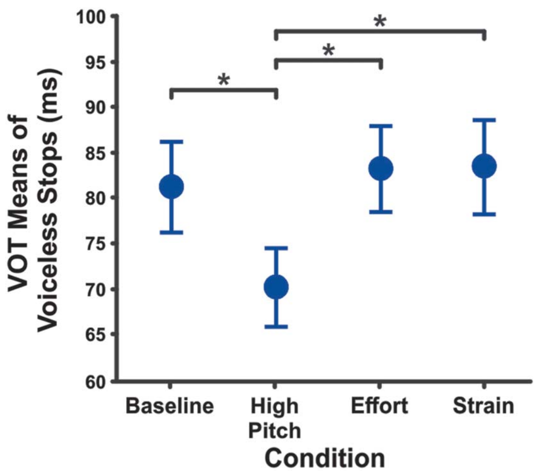
Increased vocal effort and increased vocal strain had no effect on VOT means, as shown in Figure 1, which did not support our first hypothesis. Since laryngeal tension is often associated with vocal effort and vocal strain (Lien et al., 2015; McCabe & Titze, 2002; McKenna et al., 2019; Rosenthal et al., 2014), these results seem to be in contrast

Table 2. Results of analyses of variance on voice onset time (VOT) means and standard deviations (*SD*s).

Effect	*df*	*F* value	*p* value	Effect size (η_p^2)	Qualitative effect size
VOT means					
Participant	15	20.38	< .001	.30	Large
Condition	3	12.88	< .001	.05	Small
Voicing	1	3153.40	< .001	.81	Large
Vowel	1	130.71	< .001	.15	Medium
Place of articulation	2	68.55	< .001	.16	Medium
Condition × Voicing	3	6.89	< .001	.03	Small
Condition × Vowel	3	0.38	.770	—	—
Condition × Place of Articulation	6	0.77	.600	—	—
VOT *SD*s					
Participant	15	3.31	< .001	.07	Small
Condition	3	5.08	.002	.02	Small
Voicing	1	175.49	< .001	.20	Medium
Vowel	1	0.03	.970	—	—
Place of Articulation	2	6.90	.009	.01	Small
Condition × Voicing	3	1.15	.330	—	—
Condition × Vowel	3	0.50	.810	—	—
Condition × Place of Articulation	6	0.62	.600	—	—

Note. Effect sizes and interpretations are only provided for significant effects ($p < .05$). Dashes indicate nonsignificant findings. *df* = degrees of freedom.

Figure 1. Average voice onset time (VOT) means and 95% confidence intervals of voiceless stops for each experimental condition. Brackets indicate significant differences between conditions ($p < .05$).



with the change in VOT means observed as a result of increased pitch. However, increased f_o is likely due to stiffening of the vocal folds by activation of the cricothyroid and thyroarytenoid muscles (Shipp, 1975; Stemple et al., 2018), whereas vocal effort may manifest as laryngeal tension in a number of other intrinsic and extrinsic laryngeal muscles. For example, vocal effort can be accompanied by an increase in subglottal pressure, which requires increased activation of intrinsic muscles such as the lateral cricoarytenoid and interarytenoid muscles to maintain vocal fold adduction (Chhetri & Park, 2016). Vocal effort can also occur with supraglottic compression, which is caused by the activation of muscles above the glottis (Stager et al., 2000). Thus, based on these results, it appears that individuals with typical voices engage different laryngeal musculature to increase pitch than to increase vocal effort. Alternatively, the stiffening of the vocal folds that occurs when speaking with increased vocal effort may not be a large enough increase to have a meaningful impact on VOT means. These results are in agreement with McKenna et al. (2020), which found that there was no difference in VOT means between individuals with typical voices and individuals with vocal hyperfunction.

Figure 2. Average voice onset time (VOT) standard deviations (*SD*s) and 95% confidence intervals for each experimental condition. Brackets indicate significant differences between conditions ($p < .05$).

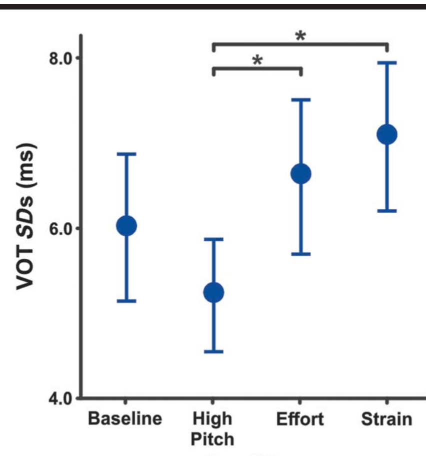


Table 3. Means (and standard deviations) of variables used to evaluate experimental fidelity for each experimental condition, including average dB SPL (SPL), average speaking rate measured as the average time per syllable in ms (Speaking rate), average increase in fundamental frequency from baseline (f_o), self-perceived vocal effort rating on a scale from 0 to 100 mm (Effort), and listener-perceived vocal strain rating on a scale from 0 to 100 mm (Strain).

	Experimental condition			
Variable	Baseline	High pitch	Effort	Strain
SPL (dB SPL)	75.6 (3.6)	85.0 (4.6)	82.2 (4.3)	80.1 (4.6)
Speaking rate (ms)	312 (31)	323 (32)	319 (32)	330 (26)
f_o (semitones)	—	7.23 (1.37)	2.12 (2.24)	1.98 (2.20)
Effort (mm)	16.9 (15.0)	42.4 (19.8)	79.1 (18.8)	81.9 (15.7)
Strain (mm)	6.2 (5.0)	11.0 (4.5)	14.0 (11.3)	35.8 (13.7)

Note. Dashes indicate that increases in fundamental frequency are not applicable to the baseline condition.

Changes in VOT SDs Across Experimental Conditions

Although condition had a significant effect on VOT *SD*s (see Table 2), there were no significant differences in VOT *SD*s between the baseline condition and the other conditions, as shown in Figure 2. These results rejected our second statistical hypothesis, but aligned with our prediction. When compared to individuals with typical voices, VOT variance has been shown to significantly increase in individuals with vocal hyperfunction, a voice disorder clinically characterized by laryngeal tension (McKenna et al., 2020). Since VOT variance increased in individuals with vocal hyperfunction in this previous study, but not in speakers with typical voices who increased laryngeal tension via increased f_o, vocal effort, and vocal strain in the current study, it appears that increased VOT variance may be intrinsic to individuals with vocal hyperfunction, irrespective of the degree of laryngeal tension currently being used by the individual. McKenna et al. suggested that increased VOT variability in individuals with vocal hyperfunction is a result of larger auditory-motor targets (McKenna et al., 2020). This argument was based on previous research that showed that individuals with vocal hyperfunction may be predisposed to auditory-motor integration deficits (Stepp et al., 2017). With this impairment in the vocal motor system, individuals with vocal hyperfunction are expected to experience larger auditory-motor targets that manifest as increased VOT *SD*s, regardless of the presence of increased laryngeal tension. Likewise, individuals without vocal hyperfunction are expected to have more consistent VOTs, regardless of the temporary use of laryngeal tension (i.e., vocal effort and strain). The results of the current study provide further support for this supposition and suggest that increased VOT *SD*s may be indicative of an auditory-motor deficit in individuals with vocal hyperfunction.

Though there were no significant differences when comparing conditions to the baseline condition, there was a significant decrease in VOT *SD*s during the high pitch condition when compared to the effort and strain conditions, as shown in Figure 2. This difference may be, in part, due to decreased VOT means during the high pitch condition. Shorter VOT means are likely to result in somewhat smaller VOT *SD*s. Additionally, during the high pitch conditions, participants were matching the frequency of a note, which often resulted in a monotonous repetition of the VOT stimuli. It is possible that restricting participants to a limited frequency range also reduced the variability of VOT durations. In contrast, during the effort and strain conditions, participants were intentionally varying their voice to result in more effortful or strained productions, which may have introduced a greater variability into acoustic measures such as VOT durations. Despite these changes, high pitch, effort, and strain were not significantly different from the baseline condition. Although differences may exist if this study were completed in a much larger group of speakers, the lack of significant differences from the baseline condition in the current study indicate that any changes in variability as a function of increased effort or strain are likely small.

Experimental Fidelity

Additional analysis was used to verify the fidelity of the experimental set up across conditions in order to ensure that the differences observed in VOT measures were not due to additional changes in voice production. For instance, vocal loudness could have an impact on VOT means. Although there is no previous work directly investigating the effect of loudness on VOT means in adults with typical voices, one study found that VOT means significantly decreased when children aged 5–12 years spoke with a loud voice (Knuttila, 2011). The authors reasoned that the increase in subglottal pressure that accompanies increased loudness causes the vocal folds to stiffen in order to maintain glottal closure. This tightening reduces the open glottal area and moves the vocal folds toward the midline of the glottis at rest, thereby resulting in faster abduction and adduction when individuals speak at an increased loudness (Holmberg et al., 1988). This, therefore, could result in a faster initiation of voicing following a voiceless consonant, ultimately resulting in shorter VOT means. Thus, the average SPL was calculated for each condition in order to determine whether there was an effect of condition on average SPL.

As shown in Table 2, the average SPL for each condition was found to be 75.6 dB SPL, 85.0 dB SPL, 82.2 dB SPL, and 80.1 dB SPL for baseline, high pitch, effort, and strain, respectively, indicating that participants spoke with increased SPL in the high pitch, effort, and strain conditions when compared to baseline. The results from previous research (Knuttila, 2011) suggest that the significant decrease in VOT means seen during the high pitch condition may be due, in part, to an increase in SPL. However, this decrease in VOT means was not seen in the effort or strain condition, despite increases in average SPL. Given that the decrease in VOT during increased pitch is well supported by previous studies (McCrea & Morris, 2005; Narayan & Bowden, 2013), it is unlikely that SPL had a meaningful effect on VOT means in this study.

Speaking rate has also been shown to have a significant effect on VOT means (Allen et al., 2003; Kessinger & Blumstein, 1997, 1998; Volaitis & Miller, 1992). Specifically, decreased speaking rate increases average syllable length, which in turn increases VOT means. In the current study, a custom-made MATLAB graphical user interface that highlighted stimuli sentence at a constant rate was used to keep a relatively constant speaking rate across participants and conditions. As shown in Table 2, the average syllable lengths for all conditions were between 312 and 330 ms. One previous study used syllable lengths of 500–799 ms for slow speech, 300–499 ms for medium speech, and 100–299 ms for fast speech and found that VOT means only increased by 23–35 ms across the three conditions (Volaitis & Miller, 1992). Given that the average syllable lengths of all conditions in the current study are well within the range for "medium speech," we concluded that speaking rate was, on average, adequately consistent.

Changes in f_o, self-ratings of vocal effort, and listener-perceptual ratings of vocal strain were calculated in order to verify that participants changed their voice in each experimental condition. From baseline to high pitch, speakers increased their f_o by an average of 7.23 ST (semitones), which confirms that speakers successfully increased their f_o during the high pitch condition. There was no meaningful difference in the average semitone increase between sexes: Male speakers increased their f_o by an average of 7.18 ST, and female speakers increased their f_o by an average of 7.28 ST. Additionally, as shown in Table 2, speakers increased their f_o by an average of 2.12 ST and 2.24 ST during the effort and strain conditions, respectively. Although speakers were instructed not to increase their pitch, it is possible that the small increases in f_o during the effort and strain conditions were due to speaking with an increased f_o during the preceding high pitch condition. Given that McCrea and Morris only observed significant differences in VOT means between low and high pitch conditions, but not between low and medium or medium and high pitch conditions, it is likely that the small increases in f_o during the effort and strain conditions did not have a meaningful effect on VOT means (McCrea & Morris, 2005).

In addition to verifying an increase in f_o during the high pitch condition, increases in self-perceived vocal effort and listener-perceived vocal strain were confirmed for the effort and strain conditions. Evaluated on a visual analog scale from 0 to 100 mm, vocal effort ratings increased from 16.9 mm at baseline to 79.1 mm during the effort condition. Using the CAPE-V visual analog scale, vocal strain ratings increased from 6.2 mm at baseline to 35.8 mm during the strain condition, indicating an increase from no strain to mild-to-moderate strain (Kempster et al., 2009). These increases in effort and strain suggest that speakers successfully increased their effort during the effort condition and their strain during the strain condition.

Differences Between Effort and Strain

Participants were instructed to produce VOT utterances using both vocal effort and vocal strain. Across all participants, the average changes in f_o from baseline (2.12 and 1.98 ST), average SPL (82.2 and 80.1 dB SPL), and average syllable duration (319 and 330 ms) were not meaningfully different between the effort and strain conditions, respectively. Although there were no significant differences in VOT means (see Figure 1) and *SD*s (see Figure 2) between the effort and strain conditions, there were differences between the participants' self-perception of vocal effort and the perceptual evaluation of strain by two voice-specializing SLPs, as shown in Table 3. Specifically, on a visual analog scale from 0 to 100 mm, participants reported an average vocal effort of 79.1 mm for the effort condition and 81.9 mm for the strain condition, indicating little difference in the degree of self-perceived vocal effort between the two conditions. In contrast, the perception of strain increased from 10.9 mm during the effort condition to 33.2 mm during the strain condition. Thus, participants perceived excessive effort during both effort and strain conditions, whereas the certified SLPs only perceived moderate strain during the strain condition. This indicates that there may be instances in which excessive effort is used by the participant, but it is not perceived as strain.

Previous work has demonstrated that self-perceived ratings of vocal effort and listener-perceptual ratings of vocal strain can have a weak correlation (Lee et al., 2005), a moderate correlation (Eadie et al., 2010), or an excellent correlation (McKenna & Stepp, 2018), depending on the study design. The variability in this relationship is likely due to the subjective nature of effort and strain assessment.

In the self-assessment of vocal effort, speakers seem to use different cues to evaluate their own voice than listeners use to evaluate external acoustic signals. One possible explanation for this is that speakers have access to additional cues such as somatosensory sensations that they can use to evaluate the presence of vocal effort, which are not present in acoustic signals (Eadie et al., 2010). This may explain why speakers tend to rate their voice as more severely dysphonic than expert listeners rate the same voice samples (Lee et al., 2005). Additionally, speakers may develop auditory and somatosensory targets based on their typical voice, which they then use for self-assessment. By becoming habituated to their own voice, individuals with voice disorders may lack the ability to fully assess their vocal function. Individuals with Parkinson's disease, for example, demonstrate a reduced ability to assess their own pitch, loudness, and overall voice quality (Kwan & Whitehill, 2011). Thus, the self-assessment of a speaker's vocal effort is dependent on their own internal framework, which may not match the framework used by expert listeners to assess vocal strain.

Individual experiences also affect listener-perceptual ratings of vocal strain. Voice assessment has been shown to vary across listeners based on the amount of training provided (Barsties et al., 2017) and the presence of anchors in the rating scale (Eadie & Kapsner-Smith, 2011). Even among expert listeners (e.g., SLPs), vocal strain has low reliability. In a study of 21 certified SLPs, strain ratings using the CAPE-V resulted in an average Pearson correlation coefficient of $r = .35$ for intrarater reliability and a Shout–Fleiss ICC coefficient of .56 for interrater reliability (Zraick et al., 2011). Based on the variability in the self-perception of vocal effort and listener-perception of vocal strain, it is unsurprising that the current study shows differences in how the effort and strain conditions are perceived by the participants and the expert listeners.

Limitations

In the current study, individuals only produced three repetitions per unique VOT utterance. This was intentional, in order to prevent potential effects of fatigue and changes to voice quality over time. However, it is possible that three repetitions may not be enough utterances to obtain consistent VOT variability measures. Future studies should investigate the effects of increased repetitions on VOT measurements.

Although participants were instructed to increase both vocal effort and vocal strain, it is possible that the participants did not meaningfully increase laryngeal tension in either condition. Laryngeal tension can be defined as the result of the activation of any combination of intrinsic and/or extrinsic muscles (McKenna et al., 2019). In the current study, it is impossible to directly determine which, if any, muscles experienced increased muscle activation. Electromyography may improve the analysis of laryngeal tension. Though electromyography is unable to capture the passive tension in muscles that results from the activation of surrounding musculature and, therefore, cannot fully detect the presence of laryngeal tension, it may help to identify the activation of targeted laryngeal muscles. Surface electromyography can be used to capture activation of extrinsic laryngeal muscles (Stepp, 2012), whereas needle electromyography can be used to capture the activation of intrinsic muscles such as the thyroarytenoid muscle (Khoddami et al., 2013). This may also help identify whether the mechanisms to increase tension as a result of increased f_o and as a result of increased vocal effort utilize the activations of different muscle groups. Future work should explore the implementation of electromyography in VOT measurements.

Lastly, individuals with typical voices who intentionally increase vocal effort may not mirror the mechanisms used by individuals with vocal hyperfunction. A lack of VOT variability in individuals using vocal effort does not explicitly indicate that VOT variability is intrinsic to individuals with vocal hyperfunction. Though previous research has used individuals with typical voices speaking with increased effort to investigate laryngeal tension and the association to vocal hyperfunction (Lien et al., 2015; McKenna et al., 2019), it is possible that individuals with typical voices do not use the same mechanisms to temporarily modulate their vocal effort as individuals who have sustained increased laryngeal tension. Future studies should investigate the VOT variability of other speakers that may exhibit sustained increased laryngeal tension, such as individuals with Parkinson's disease and occupational voice users (Goberman et al., 2002; Roy et al., 2004), and compare it to the VOT variability seen in individuals with vocal hyperfunction (McKenna et al., 2020).

Conclusions

The current study determined that VOT means shortened with increased f_o, but not with increased vocal effort or vocal strain, in young male and female speakers with typical voices. Likewise, VOT *SD*s did not change with increased f_o, increased vocal effort, or increased vocal strain when compared to baseline. These results suggest that the laryngeal tension mechanisms underlying increased f_o are different from those underlying increased vocal effort and strain. Furthermore, in conjunction with the results from McKenna et al. (2020), these results suggest that increased VOT variability may indicate a speech motor control deficit intrinsic to individuals with vocal hyperfunction.

Acknowledgments

This work was supported by the National Institutes of Health Grants R01 DC015570 (awarded to C. E. S.), P50 DC015446 (awarded to R. E. H.), and T32 DC013017 (awarded to C. A. M.) from the National Institute on Deafness and Other Communication Disorders. Thanks to Manuel Díaz Cádiz for creating custom MATLAB scripts used for data analysis and to Megan Lee for help with data analysis. Thanks to Daniel Buckley, Kimberly Dahl, and Elizabeth Heller Murray for auditory-perceptual assessment.

References

Ackermann, H., & Hertrich, I. (1997). Voice onset time in ataxic dysarthria. *Brain and Language, 56*(3), 321–333. https://doi.org/10.1006/brln.1997.1740

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *The Journal of the Acoustical Society of America, 113*(1), 544–552.https://doi.org/10.1121/1.1528172

Aronson, A. E. (1990). *Clinical voice disorders: An interdisciplinary approach* (3rd ed.). Thieme.

Auzou, P., Ozsancak, C., Morris, R. J., Jan, M., Eustache, F., & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clinical Linguistics & Phonetics, 14*(2), 131–150. https://doi.org/10.1080/026992000298878

Barsties, B., Beers, M., Ten Cate, L., Van Ballegooijen, K., Braam, L., De Groot, M., Van Der Kant, M., Kruitwagen, C., & Maryn, Y. (2017). The effect of visual feedback and training in auditory-perceptual judgment of voice quality. *Logopedics Phoniatrics Vocology, 42*(1), 1–8. https://doi.org/10.3109/14015439.2015.1091036

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot international, 5*(9–10), 341–345.

Chhetri, D. K., & Park, S. J. (2016). Interactions of subglottal pressure and neuromuscular activation on fundamental frequency and intensity. *The Laryngoscope, 126*(5), 1123–1130. https://doi.org/10.1002/lary.25550

Eadie, T. L., Kapsner, M., Rosenzweig, J., Waugh, P., Hillel, A., & Merati, A. (2010). The role of experience on judgments of dysphonia. *Journal of Voice, 24*(5), 564–573. https://doi.org/10.1016/j.jvoice.2008.12.005

Eadie, T. L., & Kapsner-Smith, M. (2011). The effect of listener experience and anchors on judgments of dysphonia. *Journal of Speech, Language, and Hearing Research, 54*(2), 430–447. https://doi.org/10.1044/1092-4388(2010/09-0205)

Gerratt, B. R., Kreiman, J., Antonanzas-Barroso, N., & Berke, G. S. (1993). Comparing internal and external standards in voice quality judgments. *Journal of Speech and Hearing Research, 36*(1), 14–20. https://doi.org/10.1044/jshr.3601.14

Goberman, A., Coelho, C., & Robb, M. (2002). Phonatory characteristics of parkinsonian speech before and after morning medication: The ON and OFF states. *Journal of Communication Disorders, 35*(3), 217–239. https://doi.org/10.1016/S0021-9924(01)00072-7

Hillman, R. E., Heaton, J. T., Masaki, A., Zeitels, S. M., & Cheyne, H. A. (2006). Ambulatory monitoring of disordered voices. *Annals of Otology, Rhinology & Laryngology, 115*(11), 795–801. https://doi.org/10.1177/000348940611501101

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*(2), 373–392. https://doi.org/10.1044/jshr.3202.373

Hoit, J. D., Solomon, N. P., & Hixon, T. J. (1993). Effect of lung volume on voice onset time (VOT). *Journal of Speech and Hearing Research, 36*(3), 516–520. https://doi.org/10.1044/jshr.3603.516

Holmberg, E. B., Hillman, R. E., & Perkell, J. S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice. *The Journal of the Acoustical Society of America, 84*(2), 511–529. https://doi.org/10.1121/1.396829

Hunter, E. J., Cantor-Cutiva, L. C., van Leer, E., van Mersbergen, M., Nanjundeswaran, C. D., Bottalico, P., Sandage, M. J., & Whitling, S. (2020). Toward a consensus description of vocal effort, vocal load, vocal loading and vocal fatigue. *Journal of Speech, Language, and Hearing Research, 63*(2), 509–532. https://doi.org/10.1044/2019_JSLHR-19-00057

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus Auditory-Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*(2), 124–132. https://doi.org/10.1044/1058-0360(2008/08-0017)

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French and English, Journal of phonetics, *25*(2), 143–168. https://doi.org/10.1006/jpho.1996.0039

Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics, 26*(2), 117–128. https://doi.org/10.1006/jpho.1997.0069

Khoddami, S. M., Ansari, N. N., Izadi, F., & Moghadam, S. T. (2013). The assessment methods of laryngeal muscle activity in muscle tension dysphonia: A review. *The Scientific World Journal, 2013*. https://doi.org/10.1155/2013/507397

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research, 18*(4), 686–706. https://doi.org/10.1044/jshr.1804.686

Knuttila, E. L. (2011). *The effects of vocal loudness and speaking rate on voice-onset time in typically developing children and children with cochlear implants* [Master's thesis, University of Alberta]. Education and Research Archive. https://doi.org/10.7939/R3K04N

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kwan, L. C., & Whitehill, T. L. (2011). Perception of speech by individuals with Parkinson's disease: A review. *Parkinson's Disease, 2011,* Article 389767. https://doi.org/10.4061/2011/389767

Lamb, K. L., Parfitt, G., & Eston, R. G. (2008). Effort perception. In N. Armstrong & W. van Mechelen (Eds.), *Paediatric exercise science and medicine* (*2nd ed.*). Oxford University Press.

Lametti, D. R., & Ostry, D. J. (2010). Postural constraints on movement variability. *Journal of Neurophysiology, 104*(2), 1061–1067. https://doi.org/10.1152/jn.00306.2010

Lane, H. L., Catania, A. C., & Stevens, S. S. (1961). Voice level: Autophonic Scale, perceived loudness and effects of sidetone. *The Journal of the Acoustical Society of America, 33*(2), 160–167. https://doi.org/10.1121/1.1908608

Lane, H., & Perkell, J. S. (2005). Control of voice-onset time in the absence of hearing. *Journal of Speech, Language, and Hearing Research, 48*(6), 1334–1343. https://doi.org/10.1044/1092-4388(2005/093)

Lee, M., Drinnan, M., & Carding, P. (2005). The reliability and validity of patient self-rating of their own voice quality. *Clinical Otolaryngology, 30*(4), 357–361. https://doi.org/10.1111/j.1365-2273.2005.01022.x

Lien, Y. A., Michener, C. M., Eadie, T. L., & Stepp, C. E. (2015). Individual monitoring of vocal effort with relative fundamental frequency: Relationships with aerodynamics and listener perception. *Journal of Speech, Language, and Hearing Research, 58*(3), 566–575. https://doi.org/10.1044/2015_JSLHR-S-14-0194

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *WORD, 20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Lowell, S. Y., Kelley, R. T., Awan, S. N., Colton, R. H., & Chan, N. H. (2012). Spectral- and cepstral-based acoustic features of dysphonic, strained voice quality. *Annals of Otology, Rhinology & Laryngology, 121*(8), 539–548. https://doi.org/10.1177/000348941212100808

McCabe, D. J., & Titze, I. R. (2002). Chant therapy for treating vocal fatigue among public school teachers. *American Journal of Speech-Language Pathology, 11*(4), 356–369. https://doi.org/10.1044/1058-0360(2002/040)

McCrea, C. R., & Morris, R. J. (2005). The effects of fundamental frequency level on voice onset time in normal adult male speakers. *Journal of Speech, Language, and Hearing Research, 48*(5), 1013–1024. https://doi.org/10.1044/1092-4388(2005/069)

McKenna, V. S., Diaz-Cadiz, M. E., Shembel, A. C., Enos, N. M., & Stepp, C. E. (2019). The relationship between physiological mechanisms and the self-perception of vocal effort. *Journal of Speech, Language, and Hearing Research, 62*(4), 815–834. https://doi.org/10.1044/2018_JSLHR-S-18-0205

McKenna, V. S., Hylkema, J. A., Tardiff, M. C., & Stepp, C. E. (2020). Voice onset time in individuals with hyperfunctional voice disorders: Evidence for disordered vocal motor control. *Journal of Speech, Language, and Hearing Research, 63*(2), 405–420. https://doi.org/10.1044/2019_JSLHR-19-00135

McKenna, V. S., & Stepp, C. E. (2018). The relationship between acoustical and perceptual measures of vocal effort. *The Journal of the Acoustical Society of America, 144*(3), 1643. https://doi.org/10.1121/1.5055234

Mehta, D. D., Van Stan, J. H., Zañartu, M., Ghassemi, M., Guttag, J. V., Espinoza, V. M., Cortes, J. P., Cheyne, H. A., & Hillman, R. E. (2015). Using ambulatory voice monitoring to investigate common voice disorders: Research update. *Frontiers in Bioengineering, 3,* 155. https://doi.org/10.3389/fbioe.2015.00155

Morris, R. J., McCrea, C. R., & Herring, K. D. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics, 36*(2), 308–317. https://doi.org/10.1016/j.wocn.2007.06.003

Narayan, C., & Bowden, M. (2013). Pitch affects voice onset time (VOT): A cross-linguistic study. *Proceedings of Meetings on Acoustics, 19*(1), 060095. https://doi.org/10.1121/1.4800681

Ohala, J. J. (1977). Speculations on pitch regulation. *Phonetica, 34*(4), 310–312. https://doi.org/10.1159/000259891

Rae, R. C. (2018). *Measures of voice onset time: A methodological study*. Bowling Green State University.

Rosenthal, A. L., Lowell, S. Y., & Colton, R. H. (2014). Aerodynamic and acoustic features of vocal effort. *Journal of Voice, 28*(2), 144–153. https://doi.org/10.1016/j.jvoice.2013.09.007

Roy, N., Merrill, R. M., Thibeault, S., Parsa, R. A., Gray, S. D., & Smith, E. M. (2004). Prevalence of voice disorders in teachers and the general population. *Journal of Speech, Language, and Hearing Research, 47*(2), 281–293. https://doi.org/10.1044/1092-4388(2004/023)

Shipp, T. (1975). Vertical laryngeal position during continuous and discrete vocal frequency change. *Journal of Speech and Hearing Research, 18*(4), 707–718. https://doi.org/10.1044/jshr.1804.707

Stager, S. V., Bielamowicz, S. A., Regnell, J. R., Gupta, A., & Barkmeier, J. M. (2000). Supraglottic activity: Evidence of vocal hyperfunction or laryngeal articulation? *Journal of Speech, Language, and Hearing Research, 43*(1), 229–238. https://doi.org/10.1044/jslhr.4301.229

Stemple, J. C., Roy, N., & Klaben, B. K. (2018). *Clinical voice pathology: Theory and management* (*6th ed.*). Plural.

Stepp, C. E. (2012). Surface electromyography for speech and swallowing systems: Measurement, analysis and interpretation, *Journal of Speech, Language, and Hearing Research, 55*(4), 1232–1246. https://doi.org/10.1044/1092-4388(2011/11-0214)

Stepp, C. E., Lester-Smith, R. A., Abur, D., Daliri, A., Noordzij, J. P., & Lupiani, A. A. (2017). Evidence for auditory-motor impairment in individuals with hyperfunctional voice disorders. *Journal of Speech, Language, and Hearing Research, 60*(6), 1545–1550. https://doi.org/10.1044/2017_JSLHR-S-16-0282

Stevens, K. N. (1977). Physics of laryngeal behavior and larynx modes. *Phonetica, 34*(4), 264–279. https://doi.org/10.1159/000259885

Swartz, B. L. (1992). Gender difference in voice onset time. *Perceptual and Motor Skills, 75*(3), 983–992. https://doi.org/10.2466/pms.1992.75.3.983

Sweeting, P. M., & Baken, R. J. (1982). Voice onset time in a normal-aged population. *Journal of Speech and Hearing Research, 25*(1), 129–134. https://doi.org/10.1044/jshr.2501.129

Titze, I. R. (1988). The physics of small-amplitude oscillation of the vocal folds. *The Journal of the Acoustical Society of America, 83*(4), 1536–1552. https://doi.org/10.1121/1.395910

Titze, I. R. (1992). Phonation threshold pressure: A missing link in glottal aerodynamics. *The Journal of the Acoustical Society of America, 91*(5), 2926–2935. https://doi.org/10.1121/1.402928

Titze, I. R., Jiang, J., & Drucker, D. G. (1988). Preliminaries to the body-cover theory of pitch control. *Journal of Voice, 1*(4), 314–319. https://doi.org/10.1016/S0892-1997(88)80004-3

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Van Den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research, 1*(3), 227–244. https://doi.org/10.1044/jshr.0103.227

Van Leer, E., & Van Mersbergen, M. (2017). Using the Borg CR10 Physical Exertion Scale to measure patient-perceived vocal effort pre and post treatment. *Journal of Voice, 31*(3), 389 e319–389 e325. https://doi.org/10.1016/j.jvoice.2016.09.023

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *The Journal of the Acoustical Society of America, 92*(2), 723–735. https://doi.org/10.1121/1.403997

Weismer, G. (1984). Articulatory characteristics of Parkinsonian dysarthria: Sepmental and phrase-level timing, spirantization, and glottal-supraglottal coordination. In M. R. McNeil, J. C. Rosenbek, & A. E. Aronson (Eds.), *The dysarthrias: Physiology, acoustics, perception, management* (pp. 101–130). College Hill.

Whiteside, S. P., Dobbin, R., & Henry, L. (2003). Patterns of variability in voice onset time: A developmental study of motor speech skills in humans. *Neuroscience Letters, 347*(1), 29–32. https://doi.org/10.1016/S0304-3940(03)00598-6

Whitfield, J. A., & Goberman, A. M. (2015). The effect of Parkinson disease on voice onset time: Temporal differences in voicing contrast. *The Journal of the Acoustical Society of America, 137*(4), 2432. https://doi.org/10.1121/1.4920874

Witte, R. S., & Witte, J. S. (2010). *Statistics*. Wiley.

Yu, V. Y., De Nil, L. F., & Pang, E. W. (2015). Effects of age, sex and syllable number on voice onset time: Evidence from children's voiceless aspirated stops. *Language and Speech, 58*(Pt. 2), 152–167. https://doi.org/10.1177/0023830914522994

Zlatin, M. A. (1974). Voicing contrast: Perceptual and productive voice onset time characteristics of adults. *The Journal of the Acoustical Society of America, 56*(3), 981–994. https://doi.org/10.1121/1.1903359

Zlatin, M. A., & Koenigsknecht, R. A. (1976). Development of the voicing contrast: A comparison of voice onset time in stop perception and production. *Journal of Speech and Hearing Research, 19*(1), 93–111. https://doi.org/10.1044/jshr.1901.93

Zraick, R. I., Kempster, G. B., Connor, N. P., Thibeault, S., Klaben, B. K., Bursac, Z., Trush, C. R., & Glaze, L. E. (2011). Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *American Journal of Speech-Language Pathology, 20*(1), 14–22. https://doi.org/10.1044/1058-0360(2010/09-0105)