Abstract: Vertical motion of the vocal folds during phonation is a possible diagnostically significant feature. However, it is difficult to judge vertical motion through the typical two-dimensional stroboscopic display. Through high-speed videomicroscopy (HSV), the dynamics of vocal fold vibration are easier to appreciate; however, the traditional HSV is also two-dimensional. Recently, a method to display a three-dimensional (3D) image of vocal fold vibration was published. This method, as well as stroboscopy and HSV, was utilized to study vertical motion magnitude and symmetry during modal and pressed phonations in normophonic speakers. Vertical motion judgments were rated as at least 16% more possible from the HSV-derived playbacks than from stroboscopy. The assessments from the 3D playback were different than those from the two-dimensional HSV playback for magnitude, however a similar trend was realized. The findings demonstrate consistently greater magnitudes of vertical motion during pressed phonations. Asymmetry of vertical motion was appreciated in both modal and pressed phonations. The results of this study concur with the concept of increased vertical motion during pressed phonation and recommend further investigations of the typicality of this important feature of vocal fold vibration in various modes and registers of normal and pathologically influenced phonation.

I. INTRODUCTION

Visualizing vocal fold vibratory behavior is widely accepted as an integral part of a complete voice evaluation. This vibratory behavior is known to move in three dimensions: laterally, longitudinally, and vertically. The lateral movement of vocal fold vibration is the most widely discussed and utilized in clinical voice evaluations as an indication of vocal fold stiffness. The longitudinal motion of vocal fold vibration has begun to be investigated and used as part of the clinical visualization protocol. However, the normal limits of variation in longitudinal motion remain unclear. The third dimension of vocal fold vibration, vertical motion, is not a common feature rated during the stroboscopic evaluation. While furthering our knowledge of lateral and longitudinal vocal fold vibratory deviations and their prevalence in various disorders is an important task, this paper narrows its scope to investigating the vertical motion of vocal fold vibration.

Vertical motion of vocal fold vibration has been suggested to have an increased magnitude during pressed or heavy phonations [1]. Hirano has related pressed phonation to be a result of the contraction of the thyroarytenoid muscle and relaxation of the vocal ligament [2]. This relation of the physiological components has been furthered to provide the concept that the relaxed vocal ligament may lead to an increase in pliable tissue, which may then be prone to move vertically during pressed phonation [1]. Conversely, less vertical motion may be appreciated during modal and, especially, falsetto phonations due to increased tension in the vocal ligament.

An increase in vertical motion during pressed phonation is a suggested contributing factor in the vocal fold pathologies of nodules and varices [1,3,4]. Pressed phonation is often realized in persons with voice disorders characterized by strain and muscle tension dysphonia. Given the relation between these common features of functional voice disorders and pressed phonation, it is natural to continue the investigation of vertical motion by studying the vibratory patterns during differing modes of phonation.

An in-depth paper on the presence and hypothetical detrimental impact of vertical motion in vocal fold vibration was accomplished [1]. The increase of vertical motion in pressed versus falsetto phonation was demonstrated, as was the intra-cycle variability of vertical motion during pressed phonation. These findings for vertical motion rely on visualization techniques, such as HSV, that provide true intra-cycle information.

Studies of the medial surface of the vocal folds in excised larynges using HSV have allowed for the observation of the vertical motion of vocal fold vibration from a view not achievable clinically [5,6]. Within such investigations, the presence of lateral, longitudinal, and vertical components of vocal fold vibration have been documented and relatively quantified. The variation of subglottal and supraglottal pressure as well as vocal fold tension has been noted to impact these components of vocal fold vibration in excised larynges.

While vertical motion appears to be a fundamental feature of vocal fold vibration, the difficulty of rating a three-dimensional behavior with intra-cycle variations from stroboscopy, a two-dimensional representation without intra-cycle information, remains. Recently, a comprehensive set of representations and image...
processing techniques to extract significant vocal fold features from HSV images was introduced [7]. Of particular interest to this paper is the vertical motion display, which allows for the observation of vertical motion from a three-dimensional (3D) playback. This technique capitalizes on the fact that the pixel intensity of the image is a quadratic function of the distance between the vocal folds, and the light source and camera lens. Thus, allowing pixel intensity to provide information regarding vertical motion. The specific implementation of the 3D display was presented at the Voice Foundation Symposium in June 2005.

The purpose of this study is to provide a preliminary investigation of the vertical motion inherent in vocal fold vibration. The research questions to achieve this goal were:

1. Can vertical motion be assessed through a three-dimensional display?
2. What is the variation in vertical motion and vertical level of approximation for normophonic speakers?
3. Does the amount of vertical motion vary with mode of phonation?

II. METHODOLOGY

Participants: Fifty-two vocally normal participants ranging in age from 18-65 years old were recruited from Columbia, SC and Charlotte, NC. Twenty-four male and twenty-eight female participants were divided among three age ranges, 18-33, 34-49, and 50-65. The data collection, storage, and use were in accordance with human subjects regulations. The data for this study was recorded at Presbyterian Hospital’s Voice Center in Charlotte, NC. The speech-language pathologists involved with data collection were specifically trained in voice and followed a specified protocol. During the process of accepting participation in the study through the informed consent form, the participants completed a short medical and voice history, as well as a modified voice informed consent form. The participants completed a short process of accepting participation in the study through the voice and followed a specified protocol. During the process of accepting participation in the study through the informed consent form, the participants completed a short medical and voice history, as well as a modified voice informed consent form. The speech-language pathologists also provided models of phonation. To achieve pressed phonation, participants were asked to phonate “as if lifting a heavy box”. The speech-language pathologists also provided models of phonation.

The HSV images were processed for motion compensation [8] and removal of reflection spots resulting in the HSV playback. Subsequently, the 3D playback movie, a multi-colored image relating to the extent of vertical motion of the vocal folds, was produced, as seen below in Fig. 1.

![Fig. 1. Three-dimensional graphic representations of a closed and open phase within a single glottal cycle.](image)

Visual Perceptual Judgments: Visual perceptual parameters were developed to assess vertical motion from the three playbacks, stroboscopy, HSV playback, and the 3D playback. Two voice scientists perceptually evaluated the dynamic visual images obtained from the fifty-two participants. The recordings of habitual phonation from the three playbacks amounted to 156 images that were judged by each perceptual rater. From HSV and 3D playbacks, 104 pressed phonation recordings were also rated. Twenty percent of the recordings were randomly introduced into the data set to obtain intra-rater reliability. Therefore, both perceptual raters judged 312 images for the features of vertical motion and vertical level of approximation. The entire data set was randomized prior to perceptual ratings.

Vertical motion was assessed for presence or absence, magnitude, and for left-to-right vocal fold symmetry of pitch and loudness allowing both intensity and frequency to be controlled for during each sample. A Kay Elemetrics Rhino-Laryngeal Stroboscopic system Model 9100B coupled to a 70-degree rigid endoscope was used. A laryngeal contact microphone was utilized to track vocal fold vibratory frequency.

High-Speed Videendoscopy: Kay Elemetrics High-Speed Video System Model 9700 equipped with a camera that captured 2,000 frames per second with 120 x 256 pixel resolution was utilized. A 70-degree rigid endoscope (Kay Elemetrics Model 9106), the same as that used in the above described procedures, and a 300 W constant Xenon light source (Kay Elemetrics Model 7152) were coupled with the system. The recording of HSV was synchronized with the acoustic recording, captured via a head-mount condenser microphone, to allow for comparisons between physiological and acoustic events. Participants were instructed to phonate the vowel /i/ at habitual pitch and during pressed phonation. To achieve pressed phonation, participants were asked to phonate “as if lifting a heavy box”. The speech-language pathologists also provided models of pressed phonation.

The HSV images were processed for motion compensation [8] and removal of reflection spots resulting in the HSV playback. Subsequently, the 3D playback movie, a multi-colored image relating to the extent of vertical motion of the vocal folds, was produced, as seen below in Fig. 1.
magnitude through stroboscopy, HSV playback, and 3D playback. **Magnitude** of vertical motion was rated, separately for the left and right vocal folds, on a six-point scale, with 0=absent, 1=severely decreased, 2=moderately decreased, 3=typical, 4=moderately increased, and 5=severely increased. **Presence** of vertical motion was understood if the magnitude was assigned a rate of 1-5. Vertical motion **symmetry** was calculated by the differences in magnitude ratings. If the ratings of the left versus right vertical motion magnitude differed, then the vertical motion magnitude was considered asymmetrical. Additionally, **vertical level of approximation** and **ability to judge** vertical motion from the images were rated categorically, as present or absent and able to judge or not able to judge, respectively.

**Statistical Analysis:** Measures from the visual-perceptual judgment of the stroboscopic, HSV, and 3D playbacks were compared. The instances and percentage of typical and atypical ratings were calculated. Correlation and paired t-tests were employed to determine intra-rater and inter-rater reliability. A correlation of above 0.70 and/or an alpha level above 0.20 on a paired t-test was considered to demonstrate a substantial reliability. An alpha level above 0.20 was utilized to determine the lack of statistically significant variation between and within the perceptual raters.

### III. RESULTS

**Presence** of vertical motion was noted bilaterally in stroboscopic, HSV, and 3D playbacks for all instances of modal and pressed phonations. No cases of unilateral or absent vertical motion were rated from the recordings of vocal fold vibration from normophonic speakers. No differences between stroboscopy, HSV, or the 3D playbacks was realized.

The **magnitude** of vertical motion was rated as typical during modal phonation for 60, 49, and 54% of playbacks for stroboscopy, HSV, and 3D playbacks, respectively. For pressed phonation, magnitude of vertical motion was less likely to be rated as typical, 34 and 42% of cases for HSV and 3D playbacks. Reduced vertical motion was realized in 15, 13, and 25% of modal phonations as visualized through stroboscopy, HSV, and the 3D playbacks. While during pressed phonations, vertical motion was appreciated to be reduced in only 10 and 14% of HSV and 3D playbacks. Increased vertical motion was apparent in 25, 38, and 21% of visualizations of modal phonation as displayed by stroboscopy, HSV, and 3D playbacks. Pressed phonations visualized through HSV and 3D playbacks were rated as having increased vertical motion in 56 and 44% of cases.

**Asymmetry** of vertical motion magnitude was noted in 22, 27, and 12% of modal phonations visualized through stroboscopy, HSV, and 3D playbacks. For pressed phonation, 22 and 14% of recordings were perceived as revealing asymmetrical magnitudes of vertical motion when viewed by HSV and 3D playbacks.

**Vertical level of approximation** was rated as unequal in 14.5, 11, and 14% of modal phonations as viewed by stroboscopic, HSV, and 3D playbacks. Similarly, for pressed phonations 11 and 14% of cases rated from HSV and 3D playbacks had perceivably unequal vertical levels of approximation.

The **ability to judge** vertical motion was calculated from each of the three playbacks. The raters reported not being able to judge vertical motion for 19% of stroboscopic files, 3% of 3D playback files, and 1% of HSV playback files. Files rated as not able to be judged were excluded from presence, magnitude, symmetry, and vertical level results.

Intra-rater reliability, as assessed by correlation and t-tests, was moderate to high for HSV playback and stroboscopy over both pressed and modal phonation, ranging from 0.52 to 0.94. For correlations below 0.70, the t-test had a p-value above 0.30 with the exception of symmetry rated from stroboscopy by judge 2. Correlation and t-tests revealed lower intra-rater reliability for judgments of magnitude from the 3D playback. Intra-rater reliability as assessed through percent agreement within one scalar level ranged from 87 to 100%, with a mean of 96.5%.

### IV. DISCUSSION

**Presence** of vertical motion was apparent throughout all evaluations of normophonic speakers. Presence was equally likely during modal and pressed phonations. Additionally, the three types of displays viewed were equally sensitive and specific to the presence of vertical motion. Given the consistency of vertical motion presence in normophonic speakers, it would be interesting to ascertain whether persons with voice disorders, especially those resulting from or resulting in decreased vocal fold mucosa pliability, demonstrate a similar consistency.

**Magnitude** of vertical motion was rated as typical for 58% of the images across all displays for modal phonation, and for 57% of the images rated from the two HSV-derived playbacks. These centralized ratings give credence to the ability to judge the vibratory feature of magnitude of vertical motion for normophonic speakers and the ability to utilize the ratings for preliminary estimates of the typicality of magnitude variations. The majority of participants exhibited increased magnitude of vertical motion during pressed phonations. However, normophonic speakers also demonstrated typical or decreased vertical motion during pressed phonation. Typical or decreased vertical motion may have been the result of achieving the vocal quality of pressed phonation by manipulating the laryngeal mechanism differently. Since increased medial compression of the vocal folds
from the contraction of the thyroarytenoid muscles would lead to an increased amount of pliable tissue, with the inclusion of the vocal ligament available to move vertically, it may be hypothesized that decreased medial compression with increased respiratory volume was utilized.

While symmetry of vertical motion magnitude has not been specifically discussed in the literature, a number of articles have discussed lateral and longitudinal asymmetries. Given the possible significance of vertical motion, assessing asymmetry allows for a more comprehensive view of vibratory behavior. The results indicate that an average of 19% of normophonic speakers exhibited asymmetry of vertical motion. The cause of variation in ratings from the 3D playback, versus the stroboscopic and HSV playback should be further explored. It is likely that the added dimension, allowing for increased accuracy when judging vertical motion, and the novelty of the 3D playback are the causes of the differences.

Vertical level of approximation was found to be unequal for at least 11% of normophonic speakers in modal and pressed phonations. A difference of 3.5% was noted between the playbacks. Vertical level was perceived as unequal more often from the 3D playback than for the HSV playback. This may be related to the additional information available from the 3D playback and the subsequent ability of the raters to use the information when making judgments of vertical level. The prevalence of unequal vertical level in normophonic speakers was unexpected.

The relatively large amount of recordings rated as not able to be judged from the stroboscopic as compared to the other playbacks is indicative of the difficulty of rating this vibratory feature through stroboscopic playbacks. This difference in ability to judge vertical motion is likely due to the fact that stroboscopy does not provide true intra-cycle information. This difficulty is highlighted by the clinical lack of reporting and utilization of vocal fold vertical motion as an indicator of laryngeal function. The widely used visual-perceptual vocal fold rating protocols include vertical level of approximation, but not vertical motion. Perhaps clinically important information is being disregarded.

V. CONCLUSION

The increased magnitude of the vertical motion of vocal fold vibration during pressed phonation for normophonic speakers strengthens the hypothesis of the detrimental impact of this type of phonation on the vocal fold tissue. There is undoubtedly additional information regarding vocal fold vibration available through the study of vertical motion. It is important to understand the typicality and variation of vertical motion for normophonic speakers as well as persons with laryngeal pathology through the clinical perspective of videomicroscopy, as well as to further investigate vertical motion using excised larynges.

Given the results of presence, magnitude, and asymmetry of vertical motion of vocal fold vibration, the clinical significance of these findings is compelling. An additional feature of vocal fold vibration that provides insight into the pliability of the vocal fold mucosa would be valuable. The finding of unequal vertical level of approximation in normophonic speakers questions the typicality of variation in vertical level. Further research to ascertain the normal limits of vertical level differences should be undertaken. Additionally, further investigation of the influence of mode and register of phonation on vertical motion should be conducted. Studying the effect of manipulating subglottal pressure during these productions in vivo will increase our knowledge of the mechanisms driving vertical motion during phonation. Since observing vertical motion is reliant on the ability to visualize intra-cycle information, it is likely that technological advancements leading to the ability to capture vocal fold vibration at higher frame rates would be beneficial. Refinement of the 3D playback to eliminate the artifacts of light reflection is also necessary.

REFERENCES