Perceptual consequences of changes in epilaryngeal area and shape

Robin A. Samlan and Jody Kreiman

Citation: Proc. Mtgs. Acoust. **19**, 060159 (2013); doi: 10.1121/1.4799525
View online: http://dx.doi.org/10.1121/1.4799525
View Table of Contents: http://asa.scitation.org/toc/pma/19/1
Published by the Acoustical Society of America

Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal

Montreal, Canada

2 - 7 June 2013

Speech Communication

Session 4aSCb: Voice and F0 Across Tasks (Poster Session)

4aSCb6. Perceptual consequences of changes in epilaryngeal area and shape

Robin A. Samlan and Jody Kreiman*

***Corresponding author's address: Division of Head and Neck Surgery, UCLA School of Medicine, 31-24 Rehab Center, Los Angeles, CA 90095-1794, jkreiman@ucla.edu**

Decreasing epilaryngeal area has been shown to increase glottal flow pulse skewing and harmonic amplitudes [Titze, JASA, 123:2733, 2008]. It is not known, however, whether listeners perceive voice quality changes when epilaryngeal area is altered, or if perceived quality is different if the area change occurs at the ventricular folds or aryepiglottic (AE) folds. In this study, a kinematic vocal tract model was used to create five epilaryngeal cavity shapes resulting from constriction and retraction of the ventricular and AE folds. Four voice sources simulating varying degrees of vocal deviation were filtered through the five shapes for a total of 20 stimuli. Fourteen listeners completed a sort and rate task. Results were analyzed using multidimensional scaling (MDS). Altering the epilaryngeal cavity shape resulted in voice quality differences, and perceptual distances differed by voice source. AE fold constriction was perceived most differently from other shapes for all talkers. Ventricular fold constriction was perceived most similar to AE constriction for 3 of the 4 voice sources. Glottal flow and acoustic differences for each epilaryngeal shape will be described and related to the perceived differences in voice quality.

Published by the Acoustical Society of America through the American Institute of Physics

Received 16 Jan 2013; published 2 Jun 2013

1. INTRODUCTION

Decreasing epilaryngeal area is the hypothetical basis of voice therapy techniques, such as "resonant voice," that are used clinically to improve voice quality for different types of voice and laryngeal disorders[1,2,3]. Decreasing epilaryngeal area increases both the skew of glottal flow pulses and harmonic amplitudes[4], yet it is not known whether listeners perceive voice quality changes when epilaryngeal area is altered, if the effect on quality is different when the area change occurs at the level of the ventricular folds versus the aryepiglottic folds, or whether the anticipated changes in the voice occur when the voice is disordered. Because such knowledge is important when deciding whether to target epilaryngeal shape change in therapy for patients with voice disorders, this study was designed to examine how modifying epilaryngeal cavity shape alters glottal flow, the acoustic signal, and voice quality for normal and disordered voices.

2. METHOD

For modeling purposes, the epilarynx was defined as a 2.4 cm segment of the vocal tract just superior to the vocal folds. A kinematic vocal tract model[5,6,7] was used to create five epilaryngeal cavity shapes, by orthogonally modifying epilaryngeal area (constricted vs. expanded, with areas of 0.2 cm^2 and 1.2 cm^2, respectively) and location of the constriction/expansion within the epilarynx (0.4 cm and 2.4 cm above the vocal folds, approximating the locations of the ventricular folds and aryepiglottic folds), all relative to a fifth "neutral" condition based on a naturally-produced vowel with epilaryngeal area equal to 0.36 cm^2 at the level of the ventricular folds and 0.57 cm^2 at the level of the aryepiglottic folds. Epilaryngeal area for the imposed segments was smoothly varied into and out of the target values. Settings for the vocal tract area function downstream from the epilarynx were identical for every simulation, and produced the vowel /a/[8].

Four sets of vocal fold model parameters were used to create normal, mildly, moderately, and severely asymmetric glottal configurations by sequentially weakening the left vocal fold (analogous to an increasingly severe vocal fold paralysis). The parameter settings used to generate each configuration are shown in Table 1. The source for each glottal configuration was then filtered through the five epilaryngeal cavity shapes and the upper vocal tract model to create an /a/ vowel. Measures of minimum and maximum glottal flow were completed from within the model script, and the cepstral peak prominence (CPP) and harmonics-to-noise ratio for 0-3500 Hz (HNR) were measured from the resultant output pressure signal using VoiceSauce[9].

TABLE (1). Settings of the kinematic vocal fold model parameters for the four glottal configurations

Glottal configuration	Adduction (cm)		Bulging (cm)		Nodal point ratio		Amplitude	Phase (radians)	
	R	*L*	*R*	*L*	*R*	*L*	*L (relative to R)*	*R*	*L*
Normal	0.08	0.08	0.1	0.1	0.8	0.8	1	0	0
Mild asymmetry	0.1	0.1	0.1	0	0.8	0.6	1	0	1.4
Moderate asymmetry	0.1	0.2	0.1	0	0.8	0.4	1	0	1.8
Severe asymmetry	0.1	0.3	0.1	0	0.8	0.4	0.5	0	1.8

To assess the perceptual impact of these manipulations, fourteen normal-hearing listeners completed a sort and rate task[10] in which the five /a/ vowels for a single glottal configuration were represented on a computer screen as differently shaped and colored icons. Listeners played the vowels by clicking each icon, and then dragged the icons to a line on the screen so that the physical distance between the icons on the line represented how different the vowels sounded (farther apart = more different in quality; vowels that sounded exactly the same were to be stacked). Each glottal configuration was presented in a different screen, so that listeners completed 4 trials, each including the five samples (neutral plus 4 epilaryngeal shape manipulations) for one glottal configuration. The order in which listeners completed the trials was randomized. Testing was combined with a second task not reported here; the combined tests lasted between 10 and 20 minutes.

To determine the perceptual importance of differences in epilaryngeal cavity shape for each glottal configuration, scores from the sort and rate task were analyzed via separate three-way (individual differences) multidimensional scaling (MDS). R^2 for the complete group of listeners was 0.57 for the normal glottal configuration, .71 for mild paralysis, .75 for moderate paralysis, and .72 for severe paralysis.

3. RESULTS

3.1 Vocal Function Analysis

Altering epilaryngeal area at the level of the aryepiglottic folds influenced glottal flow and the acoustic measures more than altering the area at the ventricular folds, but the specific effects at both locations depended on the glottal configuration (Figures 1 and 2). Aryepiglottic constriction decreased flow relative to other modeled epilaryngeal conditions, although for calculated minimum glottal flow this effect only emerged when the co-existing glottal configuration was moderately to severely asymmetric. Spectral noise levels decreased (as indicated by increasing cepstral peak prominence (CPP) and harmonics-to-noise ratio in the range 0-3500 Hz (HNR)) in the presence of an aryepiglottic constriction when the glottal configuration was normal or mildly asymmetrical, but this effect disappeared for moderate to severe asymmetries. Although the aryepiglottic expansion involved a larger change in epilaryngeal area than did the constriction, the effects on phonation were relatively small. Maximum air flow increased with increasing glottal asymmetry, and a small increase in spectral noise was observed when glottal geometry was normal to mildly asymmetric.

In contrast, much smaller functional and acoustic changes in the voice samples were associated with changes in epilaryngeal area at the level of the ventricular folds. Neither ventricular constriction nor expansion affected minimum flow. Increases in maximum flow occurred with increasing glottal asymmetry for both expansion and constriction, but these changes were quite small in the case of constriction. Finally, in the presence of ventricular constriction the HNR decreased relative to normal phonation, with effect size increasing with increasing glottal asymmetry.

Note that, because these values were obtained from modeling, all reported differences are "reliable," in the sense that identical parameter sets will always produce identical results. Thus, statistical analysis is meaningless (because there is no within-category variability), and the importance of the observed differences derives from their perceptual salience. We turn to this topic in the next section.

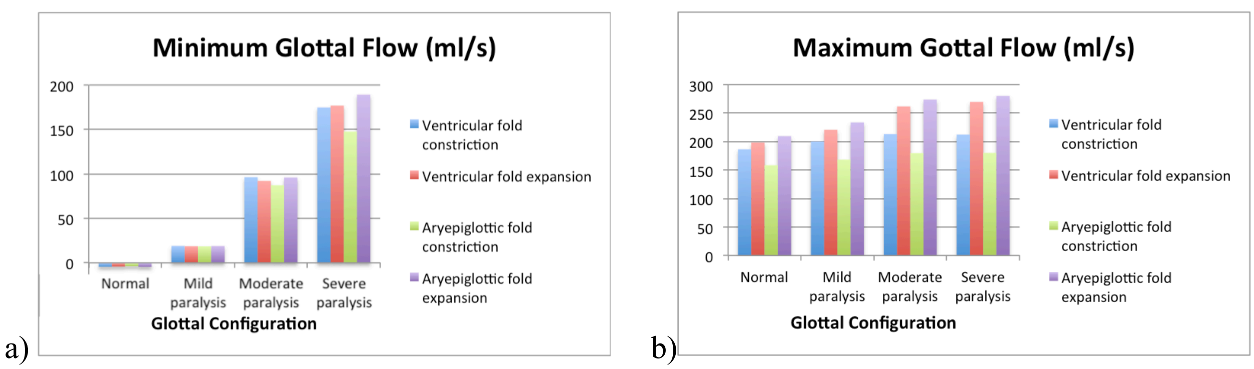

FIGURE 1: a) Minimum and b) maximum glottal flow for each epilaryngeal cavity shape for the four voice qualities.[a]

[a] Glottal flow values are not available for the baseline "neutral" configuration due to a software backup problem.

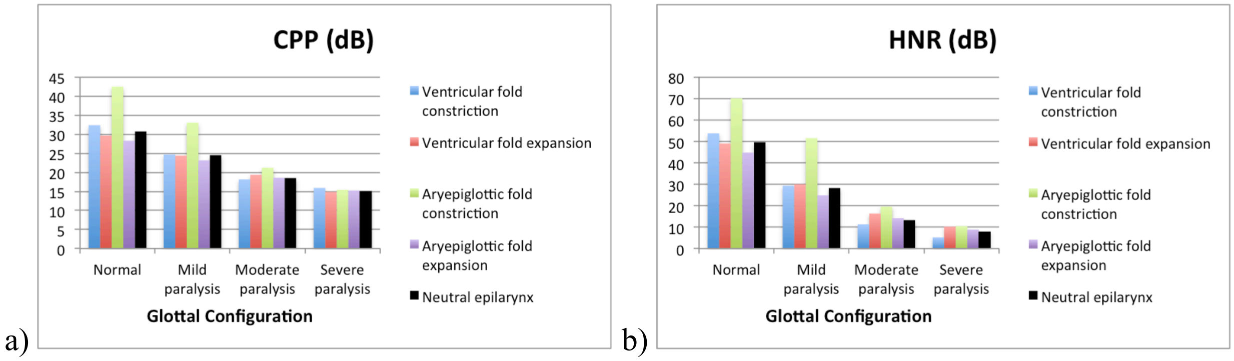

FIGURE 2: a) CPP and b) HNR for each epilaryngeal cavity shape for the five voice qualities.

3.2 Perceptual Analyses

For each glottal configuration, a one-dimensional MDS solution accounted for 57-75% of the variance in the underlying data. Figure 3 shows these solutions, for the normal, mildly, moderately, and severely paralyzed glottal configurations in panels a-d, respectively. Relative to the neutral vocal tract configuration (plotted with black circles), the perceptual importance of an epilaryngeal expansion (red and purple circles) depended on glottal configuration context, but not on the location of the expansion: Expansions were easier to hear for the normal or mildly asymmetric glottal configurations, independent of the location of the epilaryngeal expansion, and harder to hear as the overall extent of modeled paralysis worsened. In contrast, the perceptual importance of a constriction (blue and green circles) relative to the neutral configuration depended on both its location and on the glottal configuration: aryepiglottic constrictions always caused perceptible changes in quality, but ventricular constrictions had little to no perceptual impact when glottal configuration reflected a normal glottis or mild paralysis (although their perceptual importance increased for moderate to severe modeled paralyses).

From the perspective of the location of the change in epilaryngeal area, stimuli created with an aryepiglottic constriction were always perceived as very different from the neutral stimuli. In contrast, stimuli with an aryepiglottic expansion differed perceptually from neutral stimuli for normal or mildly paralyzed glottal configurations, but the perceptual importance of such expansions disappeared for moderate to severe paralyses. The perceptual impact of changes in area at the ventricular level always depended on glottal configuration: Stimuli created with a ventricular constriction were perceptually very similar to neutral stimuli when the configuration was normal to mildly asymmetric, but stimuli created with a ventricular expansion were more similar to neutral stimuli when the glottal configuration reflected a moderate or severe paralysis.

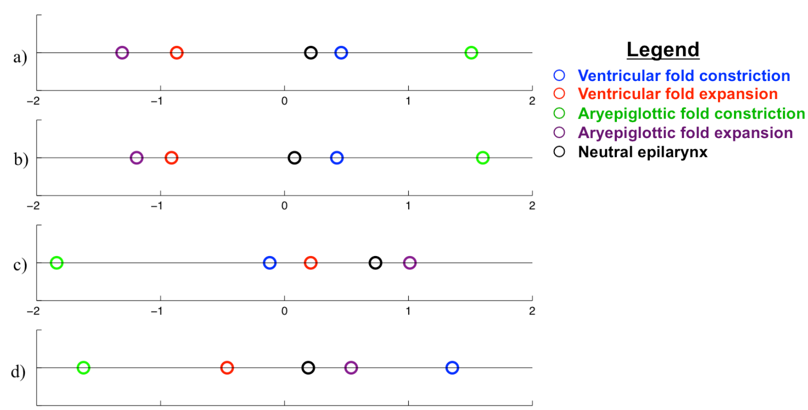

FIGURE 3: MDS solutions for a) normal, b) mild paralysis, c) moderate paralysis, and d) severe paralysis.

4. DISCUSSION

Table 2 summarizes results from these experiments. Modifying the epilaryngeal area caused perceivable differences in voice quality for normal and disordered voices. Further, listeners perceived each of the location and area modifications as having a unique quality, indicating location of constriction and degree of constriction both affected voice quality.

TABLE (2). Summary of results

Location	Area modification	Effects
Aryepiglottic	Constriction	Minimum flow ↓ for moderate/severe glottal asymmetry (relative to other modifications) Maximum flow ↓ for all glottal configurations Noise ↓ for normal/mild glottal asymmetry Large perceptual effect for all glottal configurations
	Expansion	Minimal ↑ in minimum flow for moderate/severe glottal asymmetry Maximum flow ↑ with ↑ glottal asymmetry Small ↑ in noise for normal/mild glottal asymmetry Perceptual effect for normal/mild glottal asymmetry
Ventricular	Constriction	No effect on minimum flow Limits ↑ in maximum flow with ↑ glottal asymmetry relative to expansions ↑ noise with ↑ glottal asymmetry Perceptual effect for moderate/severe glottal asymmetry
	Expansion	No effect on minimum flow ↑ maximum flow with ↑ glottal asymmetry ↑ noise for moderate/severe glottal asymmetry Perceptual effect for normal/mild glottal asymmetry

Within a particular glottal configuration, the vowel produced with aryepiglottic fold constriction was always rated most distinct from the others. This is not an indication that it sounded “better” or “worse,” because no such judgment was made during the sort and rate task. Aryepiglottic constriction limited minimum and maximum glottal flow for all voice qualities and increased CPP and HNR for normal and mildly impaired voices. Such changes are often considered to improve voice quality and are consistent with previous descriptions of increased harmonic amplitudes[4]. The elevation in CPP and HNR for the moderately impaired voice, however, was much smaller than for the milder glottal asymmetries, and the effect was not present in cases of severe asymmetry. This is presumably because the increase in noise levels associated with the glottal configuration was greater than the gain in harmonic energy contributed by the downstream constriction.

It was expected that epilaryngeal constriction would limit maximum glottal flow and enhance the harmonic structure, while expansion would increase maximum glottal flow and enhance the noise. The findings generally fit these expectations, particularly for the simulations with normal voice and mild paralysis. When the epilaryngeal area modification occurred at the level of the aryepiglottic folds, the effect was generally larger than when it occurred at the level of the ventricular folds. The exception here was that the HNR was lowest, indicating more noise, for ventricular constriction when the asymmetry was moderate and severe.

As indicated in the previous paragraphs, differences in vocal function measurements and voice quality for a particular epilaryngeal shape generally depended on glottal configuration, so that results were different when the configuration was normal/mildly asymmetric than moderately/severely asymmetric. This is likely because increased noise secondary to the simulated paralysis masked any increase in harmonic amplitude from constriction or additional noise from expansion. These findings suggest that caution is appropriate when using information about how a technique changes an essentially normal voice quality to motivate therapy for disordered voices.

Though the epilaryngeal cavity shape changes are of interest because of their suspected role as an underlying mechanism in resonant voice therapies, it is important to note that endoscopic evidence of this as an underlying mechanism is equivocal. Endoscopic findings have demonstrated aryepiglottic constriction for twang, belting, and opera singing voice qualities[11], yet constriction was not observed during resonant spoken voice productions for six speakers[12]. Additional research is needed to determine if this is the mechanism underlying resonant voice and whether that humans have access to using such a mechanism to improving speaking voice.

As a final note, the glottal area was identical within a glottal configuration, and the pharyngeal and oral cavity vocal tract area functions were held constant in this study. The differences in glottal flow, the output pressure spectrum, and voice quality were therefore fully the result of the changes to the epilaryngeal cavity shape and serve as reminder that not all changes in what we consider the “voice source” (i.e., glottal flow) should be attributed to properties of vocal fold vibration. Further study is necessary to determine whether the shapes tested improve or worsen voice quality, over what range of impairment, and whether similar perceptual differences occur for other vowels.

5. ACKNOWLEDGMENTS

Thank you to Brad Story for generously providing the model used in the study and sharing insights regarding parameter selection and data interpretation. This work was supported by research grant No. DC01797 from NIDCD, NIH.

6. REFERENCES AND LINKS

[1]I.R. Titze and B.H. Story, “Acoustic interactions of the voice source with the lower vocal tract,” J. Acoust. Soc. Am., 101, 2234-2243 (1997).
[2]I.R. Titze, “Acoustic interpretation of resonant voice,” J. Voice, 15, 519-528 (2001).
[3]I.R. Titze, “Voice training and therapy with a semi-occluded vocal tract: rationale and scientific underpinnings,” J. Speech, Lang. Hear. Res., 49, 448-459 (2006).
[4]I.R. Titze, “Nonlinear source-filter coupling in phonation: Theory,” J. Acoust. Soc. Am., 123, 2733-2749 (2008).
[5]I.R. Titze, “Parameterization of the glottal area, glottal flow, and vocal fold contact area,” J. Acoust. Soc. Am., 75, 570-580 (1984).
[6]B.H. Story, “A parametric model of the vocal tract area function for vowel and consonant simulation,” J. Acoust. Soc. Am., 117, 3231-3254 (2005).
[7]R.A. Samlan, B.H. Story, A.J. Lotto, & K. Bunton, “The acoustic and perceptual effects of left-right vocal fold asymmetries based on computational modeling” (submitted).
[8]B.H. Story, “Comparison of magnetic resonance imaging-based vocal tract area functions obtained from the same speaker in 1994 and 2002,” J. Acoust. Soc. Am., 123, 327-335 (2008).
[9]Y. Shue, P. Keating, & C. Vicenik, “VoiceSauce: a program for voice analysis,” In Proceedings of the XVIIth international congress of phonetic sciences (ICPhS), Hong Kong, 1846-1849 (2009), available for download at: http://www.ee.ucla.edu/~spapl/voicesauce/.
[10]S. Granqvist, “The visual sort and rate method for perceptual evaluation in listening tests,” Logoped. Phoniatr. Vocol. 28, 109-116 (2003).
[11]E. Yanagisawa, J. Estill, S.T. Kmucha, and S.B. Leder, “The contribution of aryepiglottic constriction to “ringing” voice quality- a videolaryngoscopic study with acoustic analysis,” J. Voice, 3, 342-350 (1989).
[12]C.G. Smith, E.M. Finnegan, and M.P. Karnell, “Resonant voice: spectral and nasoendoscopic analysis”, J. Voice, 19, 607-622 (2005).