Acta Acustica united with Acustica
Vol. 102 (2016) 209 – 213

DOI 10.3813/AAA.918937

Impact of Vocal Tract Resonance on the Perception of Voice Quality Changes Caused by Varying Vocal Fold Stiffness

Rosario Signorello, Zhaoyan Zhang, Bruce Gerratt, Jody Kreiman
UCLA School of Medicine, Head and Neck Surgery Dept., 1000 Veteran Ave. 31-24 Rehab Center, Los Angeles, CA 90095, USA. jkreiman@ucla.edu

Summary
Experiments using animal and human larynx models are often conducted without a vocal tract. While it is often assumed that the absence of a vocal tract has only small effects on vocal fold vibration, it is not actually known how sound production and quality are affected. In this study, the validity of using data obtained in the absence of a vocal tract for voice perception studies was investigated. Using a two-layer self-oscillating physical model, three series of voice stimuli were created: one produced with conditions of left-right symmetric vocal fold stiffness, and two with left-right asymmetries in vocal fold body stiffness. Each series included a set of stimuli created with a physical vocal tract, and a second set created without a physical vocal tract. Stimuli were re-synthesized to equalize the mean F0 for each series and normalized for amplitude. Listeners were asked to evaluate the three series in a sort-and-rate task. Multidimensional scaling analysis was applied to examine the perceptual interaction between the voice source and the vocal tract resonances. The results showed that the presence or absence of a vocal tract can significantly affect perception of voice quality changes due to parametric changes in vocal fold properties, except when the parametric changes in vocal fold properties produced an abrupt shift in vocal fold vibratory pattern resulting in a salient quality change.
PACS no. 43.70.-h, 43.71.-k

1. Introduction

An important goal of voice research is to establish links between voice physiology, acoustics, and perception of the produced voice. Clinically, specifying a cause and effect relationship between physiological properties of the vocal folds and voice quality perception of listeners could help surgeons and speech-language pathologists improve treatment techniques to better help improve patients' voice. Linguistically, the establishment of this relationship could lead to a better understanding of the laryngeal biomechanical adjustments that promote the differentiation of speech sounds and allow meanings to be conveyed in different languages.

Animal, human, and physical laryngeal models are often used in studies of voice production and perception. However, most of these models do not include a vocal tract. Although investigators often assume that the absence of a vocal tract may have negligible effects on vocal fold vibration in normal voice conditions [1, 2], it remains unclear how sound production and perception are affected by this simplification.

This study investigated the validity of studying voice stimuli generated by laryngeal models and obtained in the absence of a vocal tract. Although adding a vocal tract will filter the voice source and create sounds that resemble speech rather than the unnatural buzz of an unfiltered voice source, the linear source-filter theory of voice production [1] implies that source-related changes in quality from tract-free models should parallel those from models that include a vocal tract. To test this hypothesis, this study generated different series of voice stimuli from physical model experiments in which the body-layer stiffness of a physical vocal fold model was systematically varied for conditions with and without a vocal tract. Listeners were asked to evaluate each series of stimuli in a sort-and-rate experiment [3]. We hypothesized that if the presence of a vocal tract has a constant effect on voice quality variations, the perceptual scores for stimulus series produced with the same vocal fold conditions but different vocal tract conditions (with or without) should be highly correlated. A low correlation would indicate the importance of nonlinear source-tract interactions affecting voice quality perception.

Received 01 April 2015,
accepted 18 June 2015.

2. Method

2.1. Physical model experiments

The experimental setup was the same as that used in previous studies [2, 4], where it is described in detail. Briefly, the setup consisted of an expansion chamber (50.8 cm

Table I. Geometry and stiffness conditions of the physical models used in the experiments. The subscripts b and c denote the body and cover layer, respectively.

Series	I (symmetric)	II (R fold stiff body)	III (R fold soft body)
Number of steps	9	9	9
$E_{b,left}$ (kPa)	3.25–73.16	3.25–73.16	3.25–36.14
$E_{b,right}$ (kPa)	$E_{b,right} = E_{b,left}$	73.16	3.25
$E_{c,left}$ (kPa)	3.25	3.25	3.25
$E_{c,right}$ (kPa)	3.25	3.25	3.25

long, with a 23.5 cm × 25.4 cm rectangular cross section) simulating the lungs, a 11-cm straight circular PVC tube (inner diameter of 2.54 cm) simulating the trachea, and a silicon self-oscillating model of the vocal folds (described further in the following text). Two versions of this experimental setup were employed: one which included a 17-cm long vocal tract with a 2.54 cm × 2.54 cm rectangular cross section, and the second without a vocal tract.

The expansion chamber was connected upstream to a pressurized airflow supply through a 15.2-m-long rubber hose. The left and right vocal fold models were mounted on two supporting plates and were slightly compressed medially toward each other so that the glottis at rest was completely closed.

For the sake of simplicity, the two-layer physical vocal fold model had a uniform cross-sectional geometry along the anterior-posterior direction. The cross-sectional geometry was defined as in [5] and is shown in Figure 1.The vocal fold models were made by mixing a two-component liquid polymer solution (Ecoflex 0030, Smooth On, Easton, PA) with a silicone thinner solution using different composition ratios, resulting in different model stiffnesses. The Young's modulus E of each composition was measured using an indentation method [5] with a cylindrical indenter with a 1 mm diameter and an indentation depth of 1 mm on a cubic sample with dimensions 25.4 × 25.4 × 25.4 mm^3.

In the present study, for each vocal tract condition, three series of vocal folds were constructed and studied in separate experiments (see Table I). In Series I ("symmetric") the left and right vocal folds had identical geometry and material properties, and the body stiffness of both folds varied symmetrically while geometrical and cover material properties remained constant. The other two series represented left-right asymmetric conditions with varying degrees of left-right mismatch in body stiffness. These were created by varying the body stiffness of the left vocal fold while the right vocal fold remained unchanged. In series II, the right vocal fold had a very stiff body layer, while in series III, the body layer on the right was very soft.

2.2. Acoustic manipulations

Studies of voice quality require normalization of F0 and intensity, because changes in pitch and loudness can be perceptually dominant enough to preclude responses to other aspects of voice quality [6]. For this reason, the stimuli were re-synthesized to equalize the average F0 and amplitude for the appropriate stimulus series before they were used for the perceptual experiment. F0 normalization was achieved using Praat's [7] pitch-synchronous overlap-and-add (PSOLA) algorithm [8]. This approach produced signals that differed from the original in F0, but without affecting the spectrum of the voice, so that vocal quality remained the same. (For example, Kreiman *et al.* [9] found that only F0 changes of more than 40 Hz produced spectral changes, but that such changes were no greater than 1 dB.) In Series I (symmetric vocal fold stiffnesses), F0 for stimuli created without a vocal tract was adjusted to 162 Hz, and F0 for stimuli created with a vocal tract was set to 169 Hz. In Series II (asymmetric with a stiff right vocal fold body), F0 was set to 187 Hz and 185 Hz for stimuli produced with and without a vocal tract, respectively. In Series III (asymmetric with a soft right vocal fold body), F0 was adjusted to 107 Hz and 94 Hz for stimuli with and without a vocal tract, respectively. We then normalized amplitudes and downsampled the stimuli to 10 kHz through a Praat script. All stimuli were 1 second in duration.

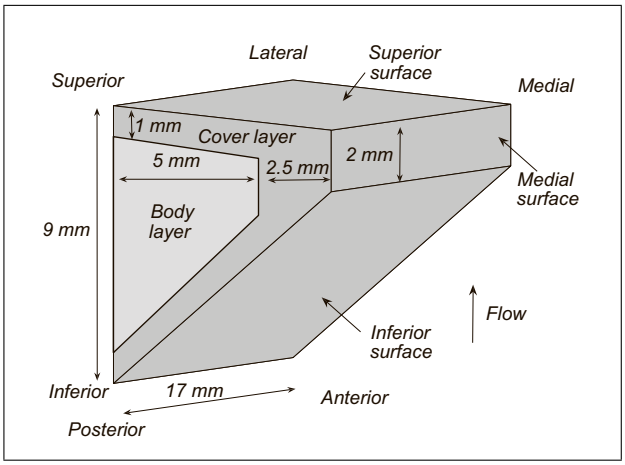

Figure 1. Representation of the physical vocal fold model used in this study.

2.3. Perceptual testing procedure

Eight listeners (3 female) took part in a perceptual test designed to evaluate the perceptual interactions between changing vocal fold stiffness and the presence/absence of a vocal tract model – in other words, the perceptual importance of source-tract interactions. None of the listeners spoke a language that uses contrastive phonation varia-

Table II. Stress and R^2 values and dimensionality chosen for representing perceptual data tests through the INDSCAL, for each series and condition.

Series	Condition	Dimensionality	Stress	R^2
I (symmetric)	with VT	3	0.19	0.33
	without VT	2	0.29	0.23
II (R fold stiff body)	with VT	4	0.14	0.47
	without VT	3	0.19	0.32
III (R fold soft body)	with VT	4	0.15	0.39
	without VT	3	0.19	0.34

Figure 2. The visual sort-and-rate task. Listeners played each stimulus by clicking an icon, then rated the similarity of the different stimuli by dragging the icons along the line such that more similarity equaled greater proximity.

tions. All reported normal hearing. Two were expert in the assessment of voice quality.

Listeners were tested individually in a double-walled sound booth. They heard stimuli at a constant comfortable listening level over Etymotic ER-1 insert earphones (Etymotic Research, Elk Grove Village, IL). They heard 6 total trials (3 symmetry conditions (Table I) × 2 vocal tract conditions), each of which included one series of 9 stimuli. Trials were presented in random order. In each trial, stimuli were presented in a visual sort-and-rate task (Figure 2; [3]). Stimuli for that trial were displayed in a random configuration as multimedia icons on a computer screen. Listeners were instructed to click the icons to play and listen to the stimuli, and then to place each icon along a straight line so that similar-sounding stimuli were close together and different-sounding stimuli were far apart. Listeners were not given any advice about the kind or extent of similarities or differences that might exist between stimuli or the nature of any underlying physical manipulations or perceptual dimensions. They were asked to put the stimuli in order according to whatever organizing percept they chose. They could listen to the stimuli as many times as they wanted, in any order. This allowed them to freely sort the series of stimuli and to adjust their sorting before confirming their response and advancing to the next trial. Although no time limits were enforced, the complete test generally lasted about 45 minutes.

Note that the procedure described above avoids the utilization of specific predetermined verbal labels to describe voice quality, such as breathy or creaky. These labels are both empirically and theoretically associated with poor listener reliability, and their use is a major source of experimental error in studies of voice quality [11].

2.4. Statistical analysis

For each listener and trial, we calculated the absolute distance between the measured placements of each pair of stimuli. Distances were assembled into dissimilarity matrices (one lower-half matrix per listener per stimulus series).These matrices were analyzed using individual differences non-metric multidimensional scaling (INDSCAL, [12]; SYSTAT software [13]), to derive normalized average distances among stimuli, to provide estimates of the perceptual dimension(s) listeners shared when they performed the sort-and-rate task, and to allow us to assess differences in perceptual patterns related to the presence vs. absence of a vocal tract.

3. Results

Results of the INDSCAL analyses are given in Table II. In all cases, the condition with a vocal tract included an extra dimension relative to the condition without a vocal tract, suggesting that adding a vocal tract added perceptual complexity to the stimuli. Given that vocal fold conditions were identical between series with and without a vocal tract, and that the same vocal tract was used in all conditions, this extra dimension can only be due to source-tract interaction, which may have induced non-uniform changes to voice quality across different vocal fold conditions, as further discussed below.

Distances between all possible pairs of stimuli in each derived multidimensional space were calculated and used in subsequent analyses. Figure 3 shows average distances for each vocal fold series and for the with- and without-vocal tract conditions. A two-way repeated measures ANOVA (within factor: vocal tract condition; between factor: vocal fold series) performed on these distances showed significant effects of presence/absence of a vocal tract ($F(1, 105) = 5.06$, $p = .027$) and vocal fold condition (symmetric, asymmetric with right fold stiff, asymmetric with right fold soft) ($F(2, 105) = 5.11$, $p = .008$), plus a significant interaction between vocal tract and vocal fold condition ($F(2, 105) = 5.92$, $p = .004$). Post-hoc Bonferroni tests indicated that the effect of vocal fold

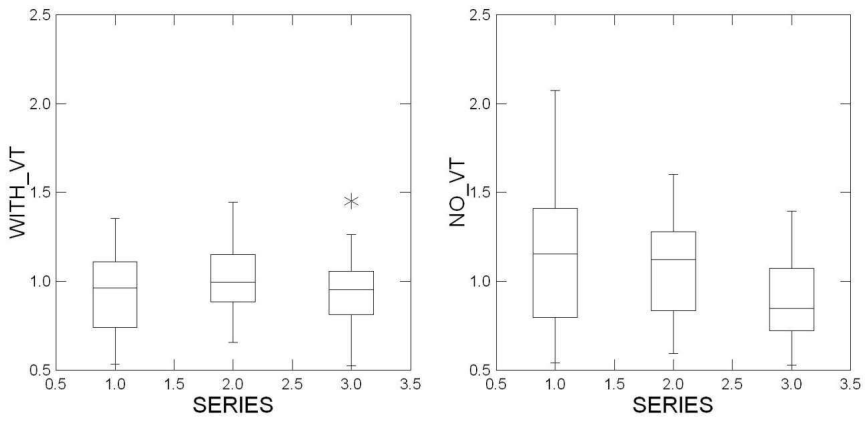

Figure 3. Box and whisker plot showing the distribution of perceptual distances for each vocal fold series, for each vocal tract condition (with and without).

condition was significant only when no vocal tract was present.

In Series I (symmetric vocal folds), a Pearson's correlation indicated that there was a weak positive relationship between the perceptual proximities in the conditions with and without vocal tract ($r = .10$, $p < .05$). This weak similarity in the perception of stimuli with and without vocal tract can be interpreted as an indication of strong source-tract interaction, so that the presence of the vocal tract caused non-uniform stiffness-dependent changes in the perception of the stimuli.

Unlike the previous series, the Pearson's correlation for Series II (asymmetric, with a stiff right vocal fold body) revealed a stronger relationship between the perceptual proximities in conditions with and without vocal tract ($r = .64$, $p < .05$). This result shows that there was a significant similarity between the perception of stimuli created with and without vocal tract conditions. As we have previously described [14], in this series changes in body stiffness of the left vocal fold model resulted in a quantal change in the vocal fold vibratory pattern, from a regime in which only the softer fold vibrated to one in which both folds vibrated together. This change in vibratory pattern resulted in a large, abrupt change in vocal quality, as a result of which listeners' responses divided stimuli into two clusters. This pattern of perceptual responses occurred for stimuli both with and without a vocal tract, suggesting that the stark differences in voice quality outweighed the perceptual contribution of the presence of the vocal tract in the context of the changes caused by the vocal fold manipulations in this condition.

In series III (asymmetric, with soft right vocal fold), similar to Series II, two regimes with distinct vibratory patterns were again observed as the body stiffness of the left vocal fold model was varied (cf. [14]). Acoustic changes associated with this change in vibratory regime were more subtle than in Series II, however, corresponding largely to changes in source spectral shape. These changes were perceptually small relative to those introduced by the presence/absence of a vocal tract: Pearson's correlation indicated that there was no relationship between the perceptual proximities in the conditions with and without vocal tract ($r = -.01$, $p < .05$).

4. Discussion and conclusions

The aim of this preliminary study was to assess the perceptual importance of interactions between the vocal tract and vocal source in voice quality changes caused by stiffness change in the vocal folds. In order to test this, we used a physical model of the vocal folds in two different conditions: with a physical vocal tract attached and without any physical vocal tract. As noted above, the validity of modeling experiments conducted without a vocal tract does not require that quality judgments be identical for stimuli created with and without a vocal tract, but merely that they be significantly linearly related. When this is the case, the contribution of the vocal tract to voice quality is constant across stimuli, so that it need not be considered when assessing the perceptual importance of changes in vocal fold configuration.

By contrast, in two of the three conditions studied here, results indicated that the perceived output of models with and without a vocal tract is not perceptually equivalent. Perceived distances between stimuli were poorly correlated, suggesting that nonlinear interactions occur between the source and filter, and that these cannot be neglected in studies of voice production. In the third condition (series II), however, listeners' judgments of stimuli produced with and without a vocal tract were significantly correlated. The major difference between series II and other series was the existence of two phonatory regimes with highly distinct vibratory patterns. In our previous study [14], the change from one vibratory regime to the other was perceptually so dominant that listeners uniformly based their perceptual judgments on this change. It is very likely that listeners in the present study similarly based their judgments on these major voice quality changes, and ignored the much more subtle changes induced by the presence of a vocal tract. Although a vibratory regime change occurred in series III, this change was smaller, as demonstrated in [14], so that it

was perceptually subordinate to the larger quality changes related to the presence of a vocal tract.

This finding – that the perceptual consequences of the presence of a vocal tract are inconsistent across experiments – makes it unsafe to generalize results from modeling studies conducted without a vocal tract to perception of phonation. This study demonstrates the importance of attaching a vocal tract in modeling experiments in which acoustic recordings are collected for perception studies of normal phonation or for which results are to be generalized to clinical situations. Perception of voice quality changes due to parametric changes in vocal fold properties can be significantly and nonlinearly affected by the presence or absence of a vocal tract, except when the parametric changes in vocal fold properties produced an abrupt shift in vocal fold vibratory pattern resulting in a salient quality change.

Acknowledgement

This study was supported by grant Nos. R01 DC011299 and R01 DC001797 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health. We thank Shaghayegh Rastifar for testing listeners.

References

[1] G. Fant: Acoustic theory of speech production. Mouton, The Hague, 1960.

[2] Z. Zhang, J. Neubauer, D. A. Berry: The influence of subglottal acoustics on laboratory models of phonation. Journal of the Acoustical Society of America **120** (2006) 1558–1569.

[3] J. Kreiman, B. R. Gerratt: Perceptual interaction of the harmonic source and noise in voice. Journal of the Acoustical Society of America **131** (2012) 492–500.

[4] Z. Zhang: Characteristics of phonation onset in a two-layer vocal fold model. Journal of the Acoustical Society of America **125** (2009) 1091–1102.

[5] D. K. Chhetri, Z. Zhang, J. Neubauer: Measurement of Young's modulus of vocal fold by indentation. Journal of Voice **25** (2011) 1–7.

[6] J. Kreiman, D. Sidtis: Foundations of voice studies: An interdisciplinary approach to voice production and perception. Wiley-Blackwell, Oxford, UK, 2011.

[7] P. Boersma, D. Weenink: Praat: doing phonetics by computer (Version 5.2.26). Computer program. Retrieved on June 14, 2011, from http://www.praat.org/, 2011.

[8] E. Moulines, F. Charpentier: Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Communication **9** (1990) 453–467.

[9] J. Kreiman, B. R. Gerratt, M. Ito: When and why listeners disagree in voice quality assessment tasks. Journal of the Acoustical Society of America **122** (2007) 2354–2364.

[10] C. M. Esposito: The effects of linguistic experience on the perception of phonation. Journal of Phonetics **38** (2010) 306–316.

[11] S. Granqvist: The visual sort and rate method for perceptual evaluation in listening tests. Logoped. Phoniatr. Vocol. **28** (2003) 109–116.

[12] J. B. Kruskal, M. Wish: Multidimensional scaling, vol. 07. Sage University Paper series on Quantitative Application in the Social Sciences, Beverly Hills and London, 1978.

[13] Cranes Software International Ltd.: Systat Software. San Jose, CA, US.

[14] Z. Zhang, J. Kreiman, B. R. Gerratt, M. Garellek: Acoustic and perceptual effects of changes in body layer stiffness in symmetric and asymmetric vocal fold models. Journal of the Acoustical Society of America **133** (2013) 453–462.