Voice Training and Therapy With a Semi-Occluded Vocal Tract: Rationale and Scientific Underpinnings

THEORETICAL/REVIEW ARTICLE

Ingo R. Titze
The University of Iowa, Iowa City, and
National Center for Voice and Speech,
The Denver Center for the Performing Arts,
Denver, CO

Purpose: Voice therapy with a semi-occluded vocal tract has a long history. The use of lip trills, tongue trills, bilabial fricatives, humming, and phonation into tubes or straws has been hailed by clinicians, singing teachers, and voice coaches as efficacious for training and rehabilitation. Little has been done, however, to provide the scientific underpinnings. The purpose of the study was to investigate the underlying physical principles behind the training and therapy approaches that use semi-occluded vocal tract shapes.

Method: Computer simulation, with a self-oscillating vocal fold model and a 44 section vocal tract, was used to elucidate source–filter interactions for lip and epilarynx tube semi-occlusions.

Results: A semi-occlusion in the front of the vocal tract (at the lips) heightens source–tract interaction by raising the mean supraglottal and intraglottal pressures. Impedance matching by vocal fold adduction and epilarynx tube narrowing can then make the voice more efficient and more economic (in terms of tissue collision).

Conclusion: The efficacious effects of a lip semi-occlusion can also be realized for nonoccluded vocal tracts by a combination of vocal fold adduction and epilarynx tube adjustments. It is reasoned that therapy approaches are designed to match the glottal impedance to the input impedance of the vocal tract.

KEY WORDS: voice therapy, voice training, singing, resonant voice, voice efficiency

Economy-oriented voice training is based on the premise that vocal injury can be minimized if vibration dose and collision stress in the vocal folds are reduced (Berry et al., 2001). One primary application is for clients who suffer from the effects of long hours of daily speaking, such as teachers in classrooms. The intent is not simply to train clients to talk softer, as in so-called "confidential voice" (Colton & Casper, 1996) or by using amplification (Roy et al., 2002), but to produce normal vocal intensity with less mechanical trauma to tissues. The current hypothesis is that increased nonlinear source–filter interaction, as in woodwind or brass musical instruments, is one way to achieve this economy. In brass instrument playing, for example, it has been shown that the lips vibrate in rather simple oscillatory motion, without much collision, in spite of abrupt pressure changes in the brass tube that help drive the lips (Ayers, 1998). By analogy, this means that the vocal tract is not only passively engaged as a filter to selectively attenuate partials of the source spectrum; rather, it is actively involved in the production of energy (in a feedback sense), allowing more aerodynamic energy to be converted into acoustic energy.

Nonlinear interaction can occur between the glottal sound source and either the subglottal tract or the supraglottal tract. Subglottal interaction is assumed to facilitate the modal register, in which the inferior portion of the vocal fold is highly involved in vibration. A strong coupling between subglottal acoustic pressures (in the form of standing waves in the trachea) and the inferior portion of the vocal fold may lead to an increase in glottal excitation (Titze, 1988). On the contrary, supraglottal interaction is assumed to facilitate the mixed register (a mixture between falsetto and modal register), in which the superior portion of the vocal fold is more involved in vibration than the inferior portion. In this case, a strong coupling between supraglottal acoustic pressures (also in the form of standing waves) and the superior portion of the vocal fold may lead to an increase in glottal excitation.

The intent of this paper is not to report training or therapy results, or specific techniques of administering therapy. Rather, the intent is to give some theoretical backing to classical approaches. For example, Aderhold (1963) described a technique of partially covering the mouth with one hand as a benefit to the actor's speaking voice, clearly a semi-occlusion of the vocal tract. Coffin (1987) described a "standing wave" exercise, where a vowel is sung while the singer covers and seals the mouth opening completely, then releases into a vocalise. Linklater (1976), a well-known theatre voice coach, adheres to the use of lip trills, an oscillatory semi-occlusion of the vocal tract, as an exercise to free the speaking voice. Singing teachers also use lip trills, tongue trills, and raspberries widely (Nix, 1999). Engel (1927) suggested that a narrowing of the mouth, formed by the tongue tip and the alveolar ridge, can produce more efficient voicing. A further development of this technique is found in Lessac (1967), who called it the "y-buzz" because the semivowel [y] creates a buzziness in the facial tissues as the acoustic pressures increase in the narrowed region of the vocal tract. Verdolini-Marston, Burke, Lessac, Glaze, and Caldwell (1995) and Verdolini, Druker, Palmer, and Samawi (1998) have built a system of voice training on the principles of energy and resonance in the speaking voice set forth by Lessac. The system was called resonant voice therapy (Verdolini, 2000). More recently, Verdolini has called it the Lessac–Madsen resonant voice training (LMRVT) method (personal communication, 2004). Some aspects of semi-occlusion of the vocal tract and the perception of resonance in the voice seem to be closely related. We will show why in this paper.

Laukkanen (1992b) and Laukkanen, Lindholm, Vilkman, Haataja, and Alku (1996) have advanced a tradition in Finland of using a bilabial voiced fricative [β:] exercise to create an ease of phonation. The bilabial voiced fricative, although not a phoneme of the English language, is rather easy to produce and satisfies the criterion of semi-occlusion in that the lips must be sufficiently approximated to produce some air turbulence. In the 1996 study, vowel production following the exercise with [β:] was produced with less muscle activity (as tested with surface electrodes) but a comparable acoustic source spectrum.

An extension of this technique is the use of flow resistant straws (Titze, 2002a; Titze, Finnegan, Laukkanen, & Jaiswal, 2002). The straws are placed between the lips and phonation occurs through them. The advantage of using a straw is that the diameter can be controlled and varied in graduated amounts by selecting from a variety of either stirring straws (small diameter) or drinking straws (large diameter). The use of artificial extensions of the vocal tract in the form of straws or tubes of various lengths and diameters has a long history (Spiess, 1899; Stein, 1937; Sovijärvi, 1964; Gundermann, 1977; Habermann, 1980; Tapani, 1992; Laukkanen, 1992a; Laukkanen, Lindholm, & Vilkman, 1995; Bele, 2005).

Other semi-occlusions of the vocal tract are the nasal consonants (/m/, /n/, or /ε/). Here the mouth is completely occluded at either the lip, alveolar, or velar position, and the velar port is opened. The nasal tract becomes the vocal tract, with the nostrils becoming the semi-occlusion. Many voice therapy techniques have been built on the frequent use of nasal consonants, including the vocal function exercises by Stemple (1993) and Stemple, Lee, D'Amico, and Pickup (1994), as well as the voice training protocols by Lessac (1967) and Verdolini (2000). Ample use of nasal consonants in the form of humming (Westerman, 1990, 1996) is a standard practice in singing training.

Collectively, lip trills, bilabial fricatives, raspberries, tongue trills, humming, and phonation into tubes or straws offer the potential for heightened interaction between the glottis and the supraglottal tract. Intraoral pressures are generally increased behind occlusions in the vocal tract, and the first formant frequency is lowered if the semi-occlusion occurs near the mouth (Bickley & Stevens, 1991). The added length of an artificial tube can further lower the first formant frequency and increase the inertive reactance of the vocal tract (Story, Laukkanen, & Titze, 2000), which heightens interactivity between the source and the filter. An equivalence of this interaction can be obtained by semi-occluding the back end of the vocal tract, that is, the epilarynx tube (Titze & Story, 1997). With these semi-occlusions (front or back), the supraglottal pressure can have a greater influence on intraglottal pressure, and therewith the airflow in the glottis (Titze, 2002b). The use of the narrowed epilarynx tube has also led to an acoustic interpretation of resonant voice (Titze, 2001) and a way of coping with unfavorable F_0–F_1 interactions when the vocal tract tends to become acoustically compliant rather than inertive (Titze, 2004a). By creating the appropriate relations between supraglottal pressure and the intraglottal pressures, a feedback nonlinearity is created that increases the maximum flow declination rate (MFDR) in the glottis (Titze, 2004b). Since MFDR is the main determinant of vocal intensity (Gauffin & Sundberg, 1989; Holmberg et al., 1988; Sapienza & Stathopoulos, 1994; Stathopoulos & Sapienza, 1997), "degree of interaction" can be used to increase intensity rather than vibrational amplitude. This is critical for maintaining vocal economy, a measure of vocal output to cost ratio (Berry et al., 2001). In the current study, a new vocal economy ratio will be examined.

As in any training program that uses exercises outside the norm of behavior, one must question transfer, or carryover, to normal behavior. Ultimately, phonation in normal speech or singing must be improved, and since speech does not maintain a constancy of semi-occluded vocal tract shapes (i.e., closed vowels, nasals, or bilabial obstruents are only intermittent), a residual effect must be demonstrated for portions of speech where the vocal tract is not occluded. This is especially the case for singers and speakers who speak loudly before audiences, using a wider-than-normal mouth opening (Appelman, 1967). Can the case be made that more source–filter interaction can be achieved with a relatively small subset of phonemes that will carry over to an overall more economic speech production? Resonance-based voice training and therapy (Laukkanen et al., 1996; Lessac, 1967; Stemple et al., 1994; Verdolini, 2000) suggests that it can. A training protocol based on an abundance of /m/, /n/, and /ɛ/ productions, and their smooth and prolonged connections to vowels, seems to improve not only normal speech, but also speech with known voice disorders (Verdolini et al., 1995). Perhaps the issue is not so much the percentage of time spent on these semi-occlusives in speech, but a change in the overall pattern of vibration that has occurred with practice. It will be shown in this paper that this change is likely to be a better impedance match between the glottis and the epilarynx tube of the vocal tract. An underlying hypothesis is that semi-occluded vocal tracts dictate a condition of vocal fold oscillation in which the vocal folds are slightly separated (abducted) by back pressure from the vocal tract. Glottal airflow and vocal fold collision are then minimized while MFDR is maximized.

In the resonant voice therapy protocols outlined by Verdolini (2000), there is a gradual acquisition of the ability to relate efficient sound production to multiple sensory modalities. Emphasis is not only on "the sound I hear" but also on "the sound I feel." The sensory experience of feeling sound in facial tissues, and the introspection thereof, is an essential part of learning to maintain economic voice production. It will be shown in this study that a semi-occluded vocal tract provides ample pressures in the oral tract (behind the occlusion) for the sensations to occur.

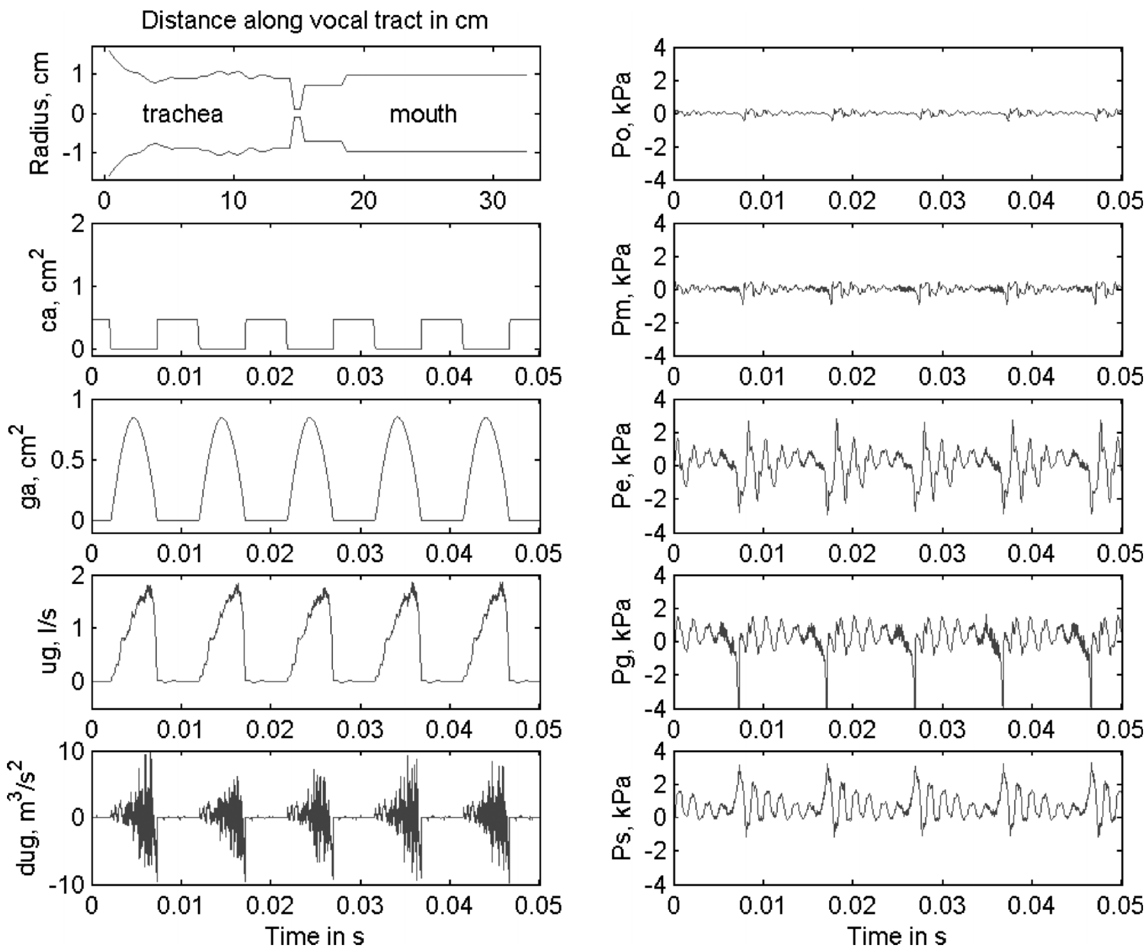
Theoretical Underpinnings

A computer simulation model was used to answer the following questions: (a) What is the relation between oral occlusion and mean intraglottal pressure? (b) As in ventriloquism, where mouth articulation can be replaced by lower vocal tract articulation, does the oral semi-occlusion have an equivalence in epilarynx tube narrowing that can be used for vowel shapes where the mouth is open? (c) Do the heightened supraglottal and intraglottal pressures reduce mean glottal flow and vibrational amplitude? (d) Is the ratio of MFDR to maximum glottal area declination rate (MADR), a new measure of vocal economy, maximized with the semi-occlusions? and (e) Is the acoustic pressure behind the lips increased so that vibratory sensations in the facial tissues are likely?

The computer model used was the three-mass body-cover model described by Story and Titze (1995) with muscle activation control as described by Titze and Story (2002). The vocal tract was simulated as a 44 section uniform tube, 3 cm^2 in cross-section and 17.5 cm in length. However, a variable diameter epilarynx tube (3.18 cm in length) replaced the first eight sections between the glottis and the pharynx to approximate the geometry of the ventrical, false fold, and laryngeal vestibule region (see Figure 1, top left graph, narrowed tube to the left of "mouth"). This reduced cross sectional area at the glottal end of the vocal tract served as an acoustic impedance matcher between the glottis and the pharynx and became an independent variable in the study. In addition, the last two sections of the vocal tract (0.8 cm in length) were used to simulate an oral semi-occlusion (to be shown later). This became the second independent variable in the study. Acoustic pressures and flows were calculated with the wave-reflection analog described in detail by Liljencrants (1985) and Story (1995). Energy losses were included as in Story (1995). Glottal flow and glottal pressures, and their interaction with vocal tract flows and pressures, were calculated as in Titze (2002b). The model also contained a 36-section subglottal (tracheal) vocal tract, which remained constant throughout the experiment (Figure 1). Future studies will address whether variable subglottal configurations can heighten interaction effects.

The biomechanical parameters of the vocal fold tissues were adjusted by muscle activation rules (Titze & Story, 2002). These rules address the manner in which vocal fold length, thickness, depth, stiffness, and

Figure 1. Computer simulation results for the wide–wide configuration. Top left is the vocal tract outline, followed by contact area (*ca*), glottal area (*ga*), glottal flow (*ug*), and glottal flow derivative (*dug*). On the right are (top to bottom) oral radiated pressure (*Po*), mouth pressure (*Pm;* directly behind the lips), pressure at the input of the epilarynx tube (*Pe*), pressure in the glottis (*Pg*), and subglottal pressure (*Ps*).

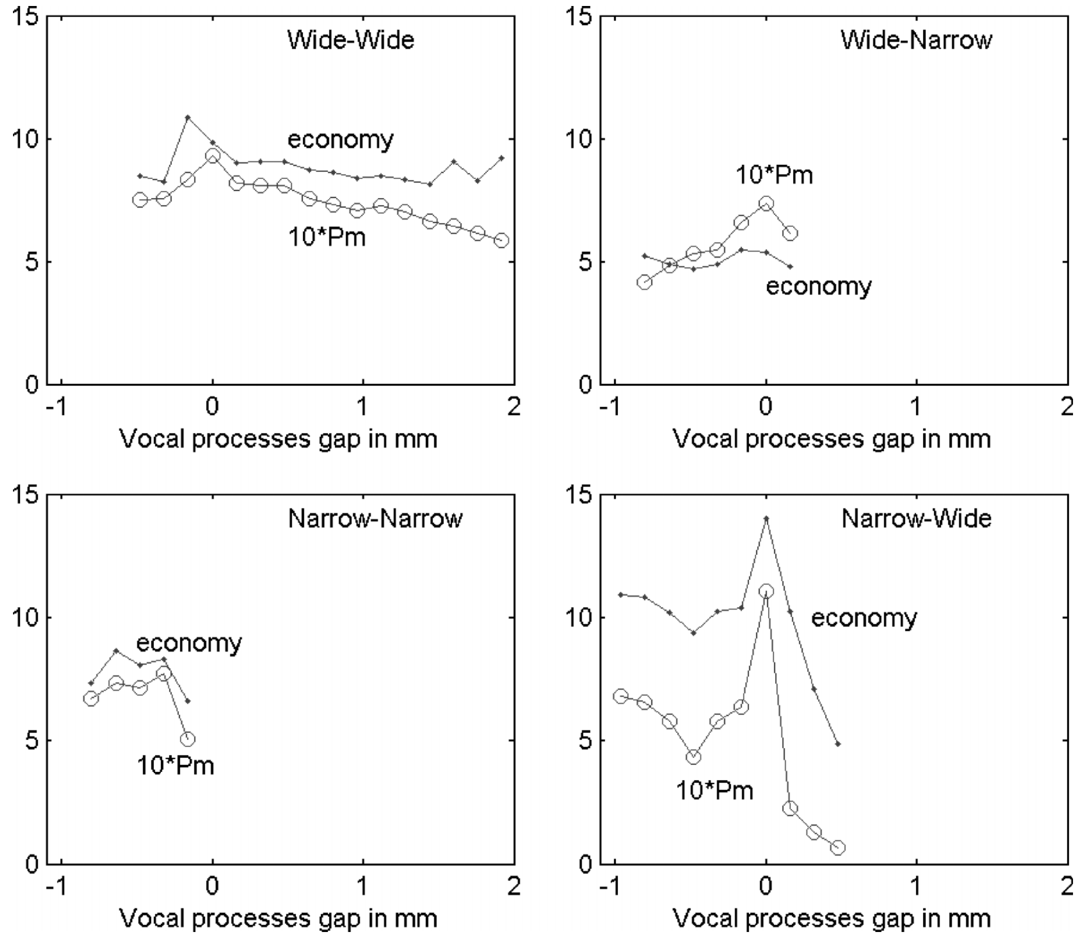


mass change with fundamental frequency and intensity, but this was not the focus of the study. For this study, both simulated cricothyroid (CT) and simulated thyroarytenoid (TA) activity were kept constant at 20% of maximum, simulated interarytenoid (IA) activity was held constant at 90% of maximum, and simulated lateral cricoarytenoid (LCA) activity was variable (the third independent variable). Simulated posterior cricoarytenoid (PCA) activity was always zero because the folds were always adducted for phonation. Alveolar (lung) pressure was also held constant at 0.8 kPa.

The model produced waveform outputs for vocal fold contact area, glottal area, glottal flow, glottal flow derivative, intraglottal pressure, subglottal pressure, supraglottal pressure, oral pressure, and lip radiated pressure. From these waveforms, the derived (dependent) variables were mean glottal flow, mean intraglottal pressure, MFDR, mean and peak glottal area, and MADR.

Waveform outputs. The first vocal tract configuration was chosen such that the mouth was nonoccluded (3.0 cm^2) and the epilarynx tube was wide (1.6 cm^2 in cross section). This configuration is henceforth referred to as the *wide–wide* configuration (see Figure 1, top left). A typical epilarynx tube cross section is 0.5 cm^2 (Story et al., 1996), making this a comparatively large (noninteractive) opening into the tract. The vocal tract impedance is low for such a wide–wide case, requiring a low glottal impedance for maximum power transfer (Titze, 2002b). This configuration may be similar to that used in "flow mode" (Sundberg, 1987) because peak glottal flows are large, as will be seen below. The glottal impedance was regulated here by vocal fold adduction, which was under control of simulated LCA activity. Self-sustained vocal fold oscillation was obtained for a range of adduction (quantified by the vocal processes gap) from −0.5 mm to 2.0 mm, with the optimum being

Figure 2. Maximization of vocal output (vocal economy and mouth pressure) by the gap between the vocal processes. The mouth pressure *Pm* (in kPa) is multiplied by 10 in order to match the scale of vocal economy.



at 0.0 mm. This optimum value was determined by calculating vocal economy (the MFDR:MADR ratio) and peak mouth pressure behind the lips for several values of adduction. Figure 2 shows these calculations. Every data point represents a condition for which self-sustained oscillation was achieved. The upper left graph is for the wide–wide case under discussion. Note the large range of vocal process gap values for sustained oscillation, and note that vocal economy and mouth pressure Pm (plotted as 10^*Pm in kPa to utilize a common scale) are maximized at a vocal process gap near 0.0 mm.

Let us return now to Figure 1. As mentioned, the top left graph shows an outline of the vocal tract configuration, including the trachea and its expansion to the left into the brochii, the vocal folds (narrowest region), the epilarynx tube (immediately to the right of the vocal folds), and the pharynx–mouth combination (labeled mouth). The combined length of the subglottal and supraglottal vocal tract was 32.5 cm, as indicated on the x-axis, and the radius is quantified along the y-axis. The remaining nine graphs are waveform outputs: contact area (ca), glottal area (ga), glottal flow (ug), glottal flow derivative (dug), oral radiated pressure (Po), mouth pressure behind the lips (Pm), epilarynx input pressure (Pe), intraglottal pressure (Pg), and subglottal pressure (Ps). The peak glottal area and peak glottal flow are very high (0.8 cm^2 and 1.8 l/s, respectively), perhaps representing an extreme case of the "flow-mode" mentioned above. The mean glottal area was 0.26 cm^2 and the mean glottal flow was 0.53 l/s. The MFDR was 7.2 m^3/s^2, as seen by the magnitude of the maximum negative spike on the lower left graph. But the flow derivative is very noisy because of the large glottal flow. In the simulation, turbulent noise was turned on in the glottis whenever the Reynolds number exceeded 1600, which it did over most of the open phase. Because the input impedance to the vocal tract was low, the mean of the epilaryngeal input pressure Pe (supraglottal pressure) was found to be near zero (0.06 kPa) and the mean of the intraglottal pressure Pg was also low (0.22 kPa). This is in relation to a mean subglottal pressure Ps of 0.6 kPa and a lung pressure of 0.8 kPa.

Figure 3. Computer simulation results for the wide–narrow configuration. Top left is the vocal tract outline, followed by contact area (*ca*), glottal area (*ga*), glottal flow (*ug*), and glottal flow derivative (*dug*). On the right are (top to bottom) oral radiated pressure (*Po*), mouth pressure (*Pm;* directly behind the lips), pressure at the input of the epilarynx tube (*Pe*), pressure in the glottis (*Pg*), and subglottal pressure (*Ps*).

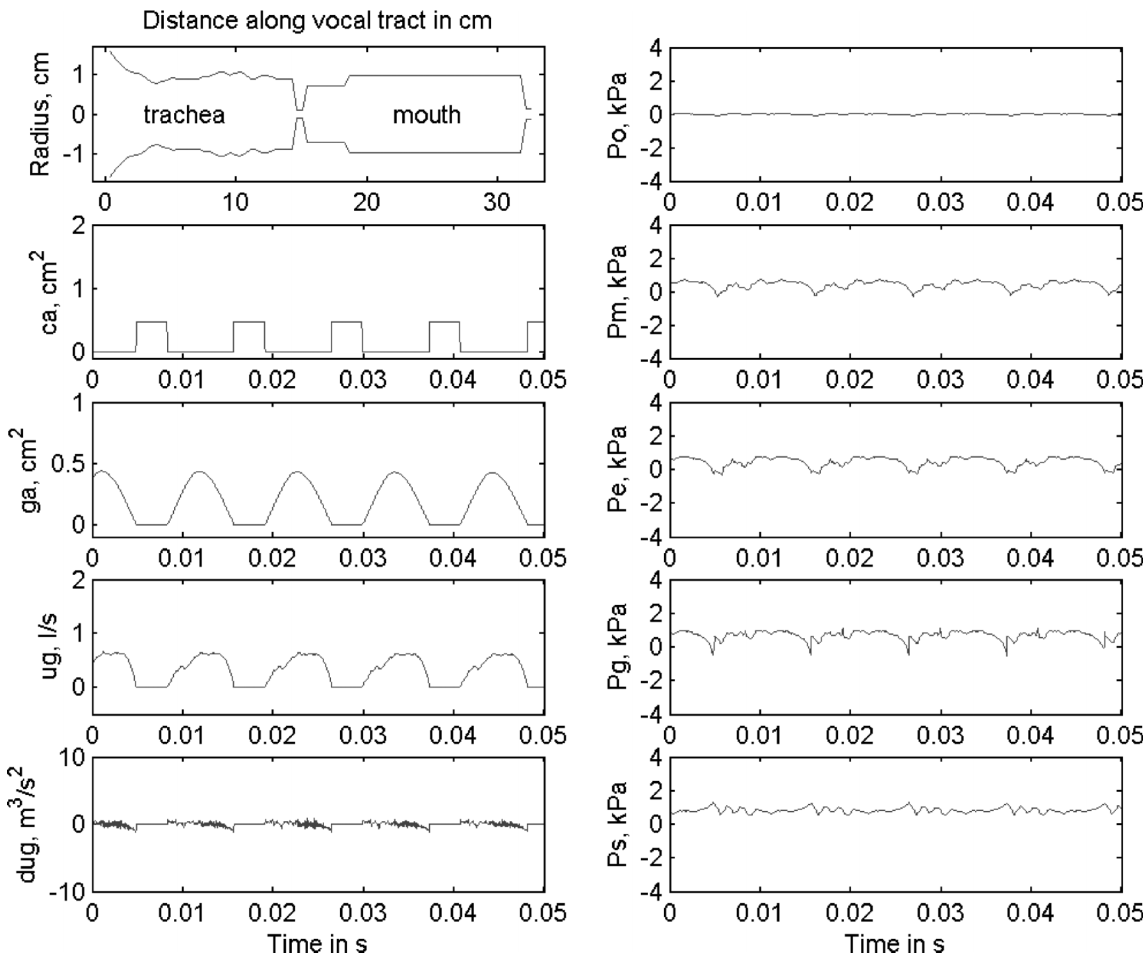
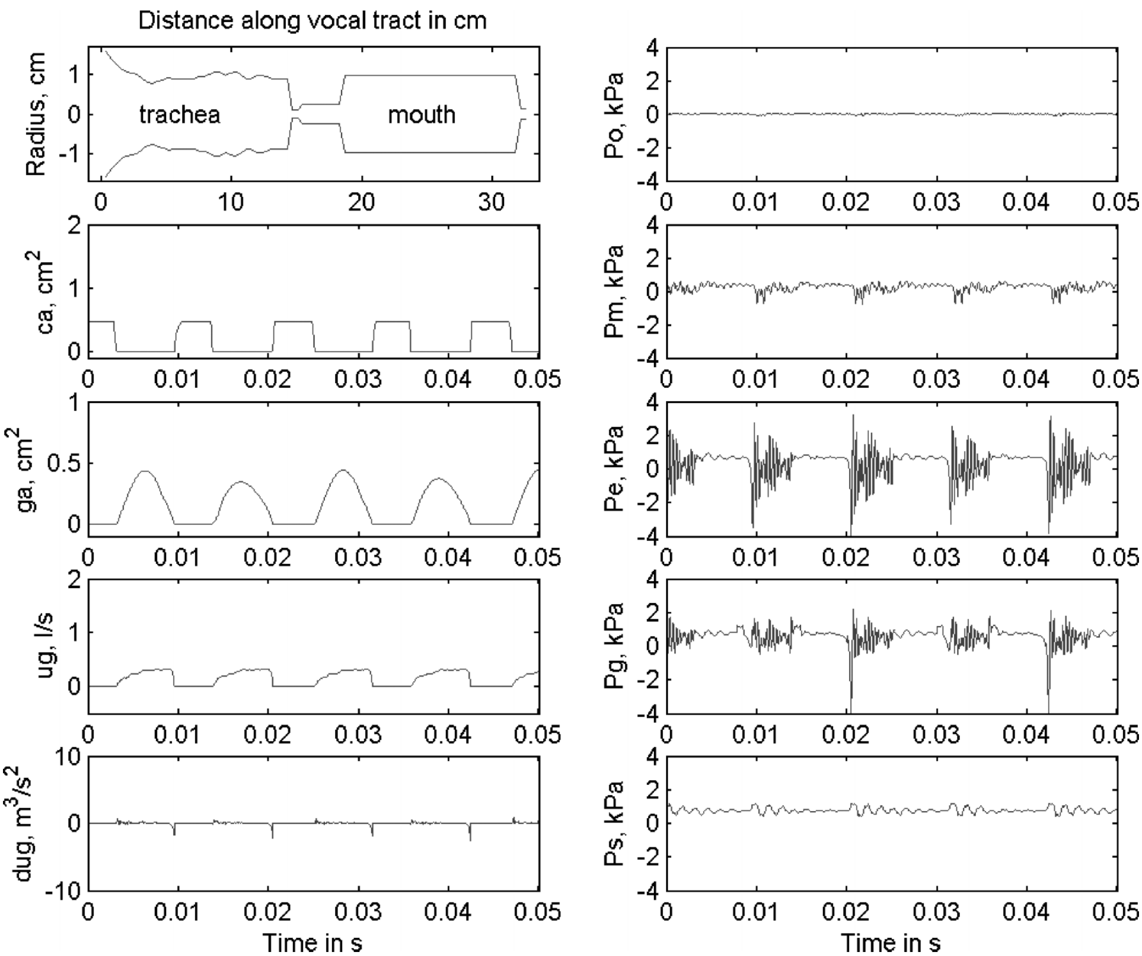


Figure 3 shows results for a *wide–narrow* vocal tract configuration. Here the epilarynx tube remained at a cross sectional area of 1.6 cm^2, but the lip area was semi-occluded to 0.05 cm^2, as in a bilabial fricative, or with the use of a small diameter stirring straw. Note that the peak glottal area was suppressed to 0.44 cm^2 and the peak glottal flow was suppressed to 0.65 l/s, both by a factor of two or more compared to the wide–wide case. The reason for this is the elevation of the mean supraglottal pressure *Pe,* which rose from 0.06 kPa to 0.40 kPa, and the concomitant elevation of the mean intraglottal pressure, which rose from 0.22 kPa to 0.51 kPa. These pressures drove the vocal folds apart slightly, resulting in a larger open quotient and an overall reduction in the peak-to-peak variation of all the acoustic pressures. The MFDR (the negative peak in the *dug* waveform) decreased from 7.2 m^3/s^2 to 1.26 m^3/s^2, as seen in the bottom left graph of Figure 2. With this dramatic reduction in glottal excitation, however, the peak mouth pressure *Pm* behind the lips decreased only slightly, from 0.93 kPa to 0.73 kPa (see Figure 3, right side, second from top; also Figure 2, top right). This suggests that the wide–narrow configuration may be useful in voice training; it minimizes glottal flow while providing the "feel" of backpressure from the vocal tract and vibration behind the lips, with virtually all of the sound being retained inside the airways. Note that the radiated pressure *Po* is very small (top right of Figure 3). Because all acoustic variations near the glottis are low, lung pressure can be raised well above normal values for speech.

Figure 4 shows results for the *narrow–narrow* vocal tract configuration. The epilarynx tube area was now very small, 0.2 cm^2, and the oral semi-occlusion was maintained at 0.05 cm^2. For optimum output, simulated

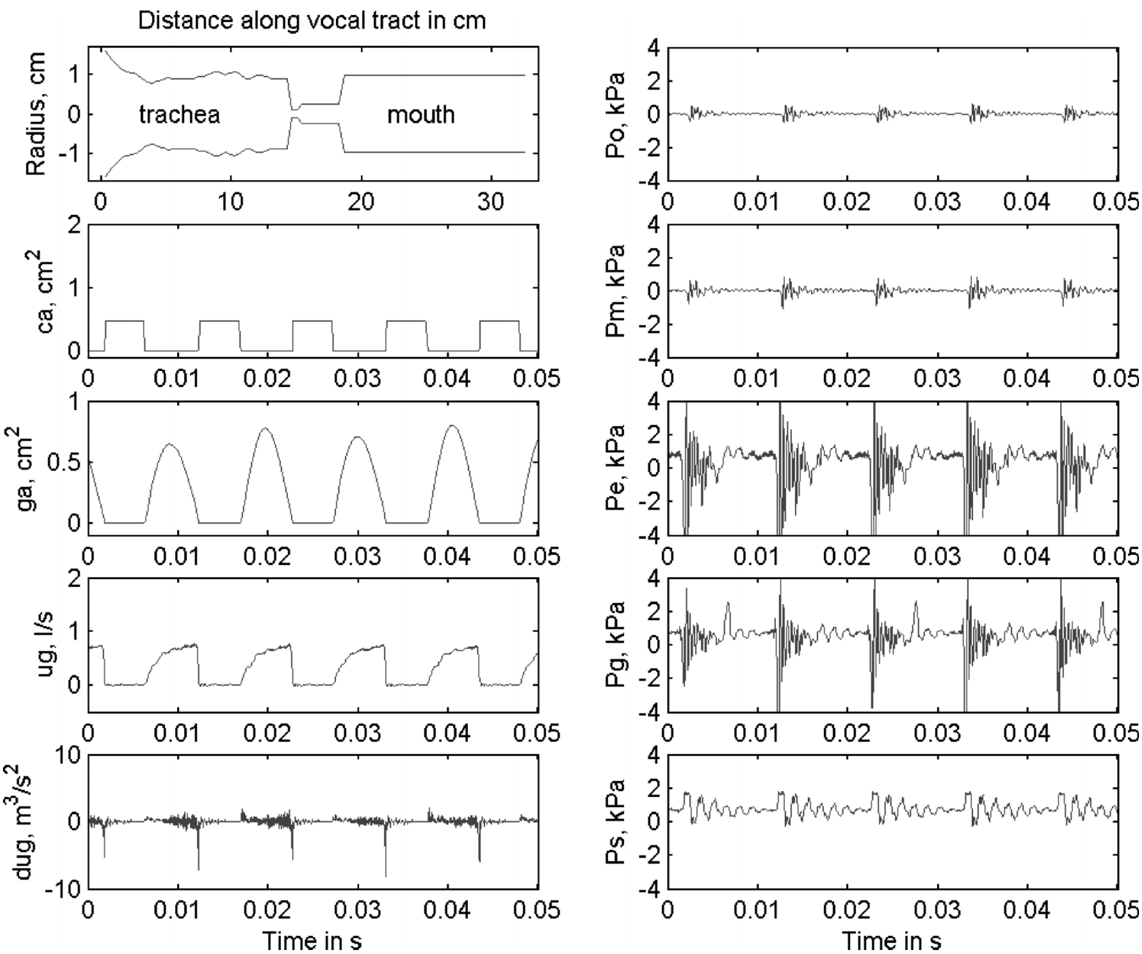
Figure 4. Computer simulation results for the narrow–narrow configuration. Top left is the vocal tract outline, followed by contact area (*ca*), glottal area (*ga*), glottal flow (*ug*), and glottal flow derivative (*dug*). On the right are (top to bottom) oral radiated pressure (*Po*), mouth pressure (*Pm*; directly behind the lips), pressure at the input of the epilarynx tube (*Pe*), pressure in the glottis (*Pg*), and subglottal pressure (*Ps*).



LCA activity was increased to 52%, which resulted in a glottal gap of −0.3 mm, that is, a prephonatory overlap of the tissue that may resemble a pressed voice in human phonation. Note that the peak glottal flow was further suppressed to 0.33 l/s and the mean glottal flow was 0.14 l/s. Waveform skewing was increased, however, resulting in a higher MFDR of 2.5 m^3/s^2. The mean supraglottal (epilaryngeal) pressure *Pe* was 0.32 kPa and the mean intraglottal pressure *Pg* was 0.55 kPa. Hence, the back pressures from the vocal tract were preserved when the epilarynx tube was narrowed. The peak mouth pressure, 0.77 kPa, was similar to the wide–narrow case. The most obvious visual differences, however, were the appearance of a period-two subharmonic in the glottal area waveform and a high frequency ring in both *Pe* and *Pg*. The subharmonic is attributed to the strong source–tract coupling and the corresponding desynchronization of vibrational modes (Mergell & Herzel, 1997). The high frequency ring is attributed to a resonance in the epilarynx tube at about 3000 Hz, known as the singer's formant (Sundberg, 1974). The reason for the resonance is that a large wave reflection takes place at the junction between the end of the epilarynx tube and the beginning of the pharynx, causing a one-quarter wavelength standing wave. A vocalist may perceive this ring as a vibratory sensation, either auditorily or vibrotactilely. Although visually the ring looks like the noise in the *dug* waveform of Figure 1, it is a completely different signal. There is no randomness in it.

Finally, Figure 5 shows simulation results for the *narrow–wide* case. The vocal tract was re-opened to 3.0 cm^2 at the mouth, but the epilarynx tube remained narrow at 0.2 cm^2. The simulated LCA activity was chosen to be 50%, again determined by optimizing the acoustic output (see Figure 2, lower right). The best glottal gap was 0.0 mm, meaning that the vocal processes were just touching. For this precise gap, this narrow–wide configuration was the most efficient for vocal

Figure 5. Computer simulation results for the narrow–wide configuration. Top left is the vocal tract outline, followed by contact area (*ca*), glottal area (*ga*), glottal flow (*ug*), and glottal flow derivative (*dug*). On the right are (top to bottom) oral radiated pressure (*Po*), mouth pressure (*Pm;* directly behind the lips), pressure at the input of the epilarynx tube (*Pe*), pressure in the glottis (*Pg*), and subglottal pressure (*Ps*).



output, yielding the greatest peak mouth pressure (1.1 kPa) and the greatest economy (14.0 cm/ms) according to Figure 2. But a small price may be paid when a vocalist tries to learn how to maintain this configuration. Small deviations in adduction on either side of this optimal 50% LCA value (0.0 mm glottal gap) caused major reductions in vocal output. In other words, the range of oscillation was more restricted and required sharper "tuning" of the glottal impedance to the vocal tract impedance. Note also that the glottal flow pulses in Figure 5 are maximally skewed, resulting in a large MFDR (8.3 m^3/s^2). This suggests that this configuration makes heavy use of vocal tract inertance, which is heightened by the narrow epilarynx tube. The peak glottal flow (0.76 l/s) was larger than for the narrow–narrow case, but not as large as for the wide–wide case. The mean glottal flow was 0.30 l/s. Back pressures were preserved, with the mean of *Pe* being 0.30 kPa and the mean of *Pg* being 0.46 kPa. Also, vocal ring was preserved and the oral radiated pressure *Po* was the largest of all four cases (top right of Figure 5). The strong source–tract interaction continued to cause some irregularities from cycle to cycle (see peaks in the *ga* waveform of Figure 5).

Derived variables. Although visual inspection of the waveforms tells much of the story, it is useful to make a few additional calculations. Table 1 shows 11 dependent variables computed from the waveforms for each of the four vocal tract configurations. Note that peak glottal flow and mean glottal flow are unusually large in the wide–wide configuration. At the computed mean glottal flow of 0.53 l/s, the entire vital capacity (4.0 l) of an average lung would be expelled in less than 8 s (at the selected 0.8 kPa lung pressure). The sound would be breathy, as was determined by the glottal noise component in Figure 1. Of course, less lung pressure could be used, but then all other variables would be scaled downward.

Table 1. Derived quantities for four vocal tract configurations.

	Wide–wide	Wide–narrow	Narrow–narrow	Narrow–wide
Peak glottal flow	1.9 l/s	0.65 l/s	0.33 l/s	0.76 l/s
Mean glottal flow	0.53 l/s	0.28 l/s	0.14 l/s	0.30 l/s
MFDR	7.2 ml/ms^2	1.26 ml/ms^2	2.5 ml/ms^2	8.3 ml/ms^2
Peak glottal area	0.85 cm^2	0.44 cm^2	0.44 cm^2	0.80 cm^2
Mean glottal area	0.26 cm^2	0.17 cm^2	0.14 cm^2	0.26 cm^2
MADR	0.74 cm^2/ms	0.23 cm^2/ms	0.30 cm^2/ms	0.60 cm^2/ms
Mean supra. pressure	0.06 kPa	0.40 kPa	0.32 kPa	0.30 kPa
Mean intra. pressure	0.22 kPa	0.51 kPa	0.55 kPa	0.46 kPa
Peak mouth pressure	0.93 kPa	0.73 kPa	0.77 kPa	1.10 kPa
Economy (MFDR/MADR)	9.8 cm/ms	5.4 cm/ms	8.3 cm/ms	14.0 cm/ms
Efficiency	0.80%	0.0019%	0.0035%	0.97%

Note. MFDR = maximum flow declination rate; MADR = maximum glottal area declination rate.

The other three configurations have more reasonable peak and mean glottal flows, with the narrow–narrow being the most conservative on expiration. With regard to MFDR (third row), the most important variable for intensity, note that the narrow–wide configuration has the largest MFDR, even though the peak flow is less than half that of the wide–wide case.

All the glottal area variables (peak glottal area, mean glottal area, and MADR) are similar between the wide–wide case and the narrow–wide case (first and last columns), and between the wide–narrow and the narrow–narrow case (second and third column), suggesting that the mouth orifice has the strongest effect on vibrational amplitudes of the vocal folds. An open mouth yields high vibration amplitudes, whereas a semi-occluded mouth yields about half the vibrational amplitudes, regardless of the epilarynx tube area. This is one reason why semi-occlusives are useful for vocal warm-up. They allow the vocalist to build up high lung pressures without excessive damage to tissues due to large vibrational amplitudes.

The back pressures from the vocal tract (i.e., the supraglottal and intraglottal pressures) are realized in all cases where there is any vocal tract narrowing (columns 2–4 in Table 1). Only the wide–wide configuration has minimal back pressure, 0.06 kPa for supraglottal pressure and 0.22 kPa for intraglottal pressure. For the other configurations, mean supraglottal pressures were on the order of 0.3–0.4 kPa and mean intraglottal pressures were on the order of 0.5 kPa, *which is more than half the applied lung pressure*. These pressures tend to keep the vocal folds separated, usually requiring a little more adductive force to match glottal and vocal tract impedances and maintain maximum power transfer.

The ninth variable in the table is a vocal economy index, newly defined and computed as the MFDR:MADR ratio. Its units are (cm^3/ms^2)/(cm^2/ms), or cm/ms, which are the units of velocity. A large MFDR is desirable for high vocal intensity, but a small MADR is desirable for conserving vibrational amplitude and maintaining small vocal fold collision velocity. MADR is proportional to the maximum tissue velocity during glottal closing, which usually occurs right before impact. In theory, this quantity should be minimized. Hence, it occurs in the denominator of our vocal economy ratio. Note in Table 1 that the wide–narrow has the poorest economy (a value of 5.4 cm/ms), the wide–wide and narrow–narrow have similar values of economy (9.8 cm/ms and 8.3 cm/ms, respectively), and the narrow–wide has the best economy (14.0 cm/ms).

Finally, the last row in Table 1 shows the traditional vocal efficiency calculation (Schutte, 1980, 1984). This is a ratio of oral radiated power (in watts) to mean aerodynamic power (mean subglottal pressure times mean glottal flow). Since most of the acoustic power is not radiated from the mouth, but rather reflected back into the vocal tract and ultimately dissipated, this measure of vocal efficiency is a small number, usually less than 1%. Note that for the four cases under consideration in Table 1, the narrow–wide tube is most efficient (0.97%), followed by the wide–wide tube (0.8%). The two cases with oral semi-occlusions are very inefficient because little acoustic power is radiated from the semi-occluded mouth. The traditional vocal efficiency measure is therefore not a good measure for assessing glottal efficiency because it is so sensitive to mouth opening.

Conclusions from simulation. Overall, the narrow-wide vocal tract (narrow at the epilarynx tube and wide at the lips) is the preferred configuration for maximized vocal output. It resembles a trumpet or megaphone shape. When properly impedance-matched, it has the highest efficiency, the highest economy, moderate peak and mean glottal flows, the largest MFDR, and it maintains a back pressure in the vocal tract near the glottis. Its only drawback is a relatively high vibrational amplitude, as measured by the peak glottal area (0.80 cm^2). Because well tuned adduction is necessary, this configuration is not always supported in self-sustained oscillation. The vocalist must learn to regulate the position of the vocal processes to a "just touching" or "almost touching" state under changing F_0 intensity and vowel conditions. This involves training of the intrinsic laryngeal muscles, especially over a large range of fundamental frequencies. Given that all five intrinsic laryngeal muscles usually cocontract differently at different fundamental frequencies (e.g., Hirano, Vennard, & Ohala, 1970), the learning process may not be trivial. For singers, this coordination of muscles is often improved with staccato and messa di voce exercises, for which rapid and precise vocal fold posturing is needed. The wide–narrow configuration, on the other hand, is the most forgiving. It requires less adductory tuning and maintains low acoustic pressures and lower vibrational amplitudes at the glottis, but retains the acoustic pressure behind the lips. This is perhaps why this inverted trumpet (or inverted megaphone) shape is desirable as a starting point for training and therapy.

It is necessary to point out that although the four configurations described here are extreme configurations, they are not necessarily unrealistic. The epilarynx tube cross sectional area is usually larger than 0.2 cm^2 and smaller than 1.6 cm^2. Also, the 0.05 cm^2 oral occlusion is a severe case, smaller than what would be experienced in a lip trill, humming, or with resonance tubes or drinking straws. But it is appropriate for a stirring straw or a bilabial fricative. Thus, the effects may be more subtle in some protocols designed for human phonation.

It is also necessary to point out that the narrow–wide configuration is a relative configuration, rather than an absolute one. Classical singers and some orators prefer to widen the pharynx and the oral cavity (Appelman, 1967). This "open throat" configuration may be in lieu of narrowing the epilarynx tube, thus preserving the ratio between the cross sectional areas for vocal ring. A lower glottal impedance (less adduction) can then balance the overall wider configuration, and all of the airflows will be increased. Thus, a vocalist can learn the trade-offs between low pressure combined with high flows and high pressure combined with low flows.

Finally, adduction training to match impedances may involve the entire medial surface of the vocal fold. Regulation of glottal impedance may involve a combination of LCA, TA, and IA muscles rather than LCA alone. Future modeling studies must include shaping parameters of the medial surface of the vocal folds in identifying an ideal glottal configuration.

Rationale for a Therapy Protocol

Semi-occluded vocal tracts are beneficial for voice therapy because they heighten interaction between the source and the filter. Such interaction can increase vocal intensity, efficiency, and economy. But for the interaction to be maximally useful, some impedance matching needs to take place between the source and the filter. This is done with a combination of glottal adduction (or abduction) and epilaryngeal tube narrowing (or widening). To get the vocalist aware of this impedance-matching requirement, some internal sensations are helpful. First is the sensation of a back pressure, perhaps the perception of a light resistance to sound emission from the vocal tract. This is the purpose of the semi-occlusion, which in therapy begins in the front of the vocal tract because of its ease of control and execution. A comfort level must be achieved in using large lung pressures over large pitch ranges while only a small amount of sound is being emitted from a small orifice. In an effort to become more efficient with this perception of resistance to sound release, there is a likelihood that some epilaryngeal tube narrowing will occur when the mouth opens to preserve this sensation. In other words, as in ventriloquism, a front articulation is replaced or augmented by a back articulation.

In addition, the vocal folds may change shape (from a convergent glottal shape where there is tight adduction of the vocal processes but loose adduction at the bottom of the folds, to a rectangular glottal shape where there is equal adduction top-to-bottom). This lowers the phonation threshold pressure. Contraction of the thyroarytenoid muscle facilitates this "squaring up" of the vocal folds. Thus, hyperadduction at the vocal processes is traded for less (but more uniform) adduction throughout the glottis. In muscle activation terms, LCA activity is traded for a little more TA activity.

More efficient sound production then leads to another sensation, one of tissue vibration all over the facial structures. This is the result of heightened acoustic pressures in the mouth (recall that there were large mouth pressures even when peak glottal flow was drastically reduced in the semi-occluded cases). As the mouth is gradually opened, a vocal ring may also be detected (usually more aurally than through tissue vibration), which is the confirmation of an epilaryngeal tube resonator. Thus, in the process of progressing from wide–narrow to narrow–narrow to narrow–wide in training, the epilarynx tube shape and the glottis are adjusted so that the economizing effects of a front occlusion are transferred to the back of the vocal tract. The oral cavity can then be adjusted for any vowel or consonant.

In an ideal therapy or voice training regime, one could rationalize that semi-occlusives be used in the order "greatest effect, but most artificial" to "smallest effect, but closest to natural." This rationale would suggest the following order:

1. Highly resistant (small diameter) stirring straw
2. Less resistant (larger diameter) drinking straw

3. Bilabial or labiodental voiced fricative
4. Lip or tongue trill
5. Nasal consonants
6. Vowels /u/ and /i/

But owing to the unfamiliarity of some clients with producing sound with a nearly closed mouth, clinicians and trainers sometimes have to start in the middle of the order, perhaps with humming or lip trilling, or even with the vowels /u/ and /i/. This prevents the vocalist from pushing, choking, or otherwise straining. As the "feel" of the production becomes more familiar and comfortable, the order given above may lead to the fastest results. Thus, the Stemple et al. (1994) vocal function exercises, or the Verdolini (2000) resonant voice exercises, may benefit from a structuring that progresses from greater degree of occlusion to lesser degree of occlusion.

Nonspeech exercises usually consist of repeated pitch glides, gradually increasing the frequency range until two octaves or more can be produced easily. But, again, initial ranges may be no more than a fifth of an octave. To create more practice variety, the melody of a simple song can be executed through the semi-occlusion. For immediate carryover into speech, the intonation and stress patterns of a spoken sentence may also be phonated through the tube or other vocal tract occlusion. Only small amplitude vibration will occur, as has been shown (Titze, Finnegan, Laukkanen, & Jaiswal, 2002); hence, there is little concern about damage to the vocal folds. Usually, after adequate familiarity with the feel of a slight resistance to sound emission, voice registers disappear because edge-vibration of the vocal folds is facilitated. Lung pressures can safely be taken up to large values without concern for injury. This has the added benefit of warming-up the respiratory muscles (in the context of breath support) without taxing the vocal folds.

It is surmised that of all the six progressive exercises listed above, compliance for out-of-therapy-room practice is most easily achievable with straw phonation. It produces the least amount of sound, thereby drawing little attention to itself. Exercises can be done in the car, walking on the street, and in hotel rooms. Most importantly, the sounds are not interpreted as speech sounds by standby listeners; hence, relatively little attention is paid to them. The desired effect of training source–filter interaction is accomplished in the least amount of time.

Acknowledgments

Funding for this work was provided by National Institute on Deafness and Other Communication Disorders Grant R01 DC04224-05.

References

Aderhold, E. (1963). Sprecherziehung des Schauspielers [Speech training of the actor]. *Grundlagen und Methoden* [*Principles and methods*]. Berlin: Henschelverlag.

Appelman, D. R. (1967). *The science of vocal pedagogy: Theory and application*. Bloomington: Indiana University Press.

Ayers, D. (1998). Observation of the brass player's lips in motion. *Journal of the Acoustical Society of America, 103,* 2873–2874.

Bele, I. V. (2005). Artificially lengthened and constricted vocal tract in vocal training methods. *Logopedics Phoniatrics Vocology, 30,* 34–40.

Berry, D., Verdolini, K., Montequin, D. W., Hess, M. M., Chan, R. W., & Titze, I. R. (2001). A quantitative output-cost ratio in voice production. *Journal of Speech, Language, and Hearing Research, 44,* 29–37.

Bickley, C. A., & Stevens, K. N. (1991). Effects of vocal tract constriction on the glottal source: Data from voiced consonants. In T. Baer, C. Sasaki, & K. Harris (Eds.), *Vocal fold physiology: Laryngeal function in phonation and respiration* (pp. 239–253). San Diego, CA: Singular.

Coffin, B. (1987). *Sounds of Singing* (2nd ed.). Metuchen, NJ: Scarecrow.

Colton, R. H., & Casper, J. (1996). *Understanding voice problems: A physiological perspective for diagnosis and treatment* (2nd ed.). Baltimore: Williams & Wilkins.

Engel, E. (1927). *Stimmbildungslehre* [*Voice pedagogy*]. Dresden, Germany: Weise.

Gauffin, J., & Sundberg, J. (1989). Spectral correlates of glottal voice source waveform characteristics. *Journal of Speech and Hearing Research, 32,* 556–565.

Gundermann, H. (1977). *Die Behandlung der gestorten Sprechstimme* [*The treatment of the pathological speaking voice*]. Stuttgart, Germany: Fischer.

Habermann, G. (1980). Funktionelle Stimmstörungen und ihre Behandlung [Functional voice disorders and their treatment]. *Archives of Oto-Rhino-Laryngology, 227,* 171–345.

Hirano, M., Vennard, W., & Ohala, J. (1970). Regulation of register, pitch, and intensity in voice. *Folia Phoniatrica, 22,* 1–20.

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. *Journal of the Acoustical Society of America, 84,* 511–529.

Laukkanen, A-M. (1992a). About the so-called "resonance tubes" used in Finnish voice training practice. *Scandinavian Journal of Logopedics and Phoniatrics, 17,* 151–161.

Laukkanen, A-M. (1992b). Voiced bilabial fricative [β:] as a vocal exercise. *Scandinavian Journal of Logopedics and Phoniatrics, 17,* 181–189.

Laukkanen, A-M., Lindholm, P., & Vilkman, E. (1995). Phonation into a tube as a voice training method. Acoustic and physiologic observations. *Folia Phoniatrica et Logopaedica, 47,* 331–338.

Laukkanen, A-M., Lindholm, P., Vilkman, E., Haataja, K., & Alku, P. (1996). A physiological and acoustic study on voiced bilabial fricative [β:] as a vocal exercise. *Journal of Voice, 10,* 67–77.

Lessac, A. (1967). *The use and training of the human voice* (2nd ed.). Mountain View, CA: Mayfield.

Liljencrants, J. (1985). *Speech synthesis with a reflection-type line analog*. Stockholm, Sweden: Royal Institute of Technology.

Linklater, K. (1976). *Freeing the natural voice*. New York: Drama Book.

Mergell, P., & Herzel, H. (1997). Modelling biphonation: The role of the vocal tract. *Speech Communication, 22,* 141–154.

Nix, J. (1999). Lip trills and raspberries: "High spit factor" alternatives to the nasal continuant consonants. *Journal of Singing, 55,* 15–19.

Roy, N., Weinrich, B., Gray, S. D., Tanner, K., & Toledo, S. W., Dove, H., et al. (2002). Voice amplification versus vocal hygiene instruction for teachers with voice disorders: A treatment outcomes study. *Journal of Speech, Language, and Hearing Research, 45,* 625–638.

Sapienza, C., & Stathopoulos, E. (1994). Comparison of maximum flow declination rate: Children versus adults. *Journal of Voice, 8,* 240–247.

Schutte, H. (1980). *The efficiency of voice production*. Groningen, The Netherlands: State University Hospital.

Schutte, H. (1984). Efficiency of professional singing voices in terms of energy ratio. *Folia Phoniatrica, 36,* 267–272.

Sovijärvi, A. (1964). Die Bestimmung der Stimmkategorien mittels Resonanzröhren [Determination of voice categories with resonance tubes]. *International Kongress Phoniatric Wissenschaft, 5,* 532–535.

Spiess, G. (1899). Methodische Behandlung der nervösen Aphonie und einiger anderer Stimmstörungen [Methodological treatment of neurologic aphonia and several other voice disorders]. *Archives of Laryngology and Rhinology, 9,* 368–376.

Stathopoulous, E., & Sapienza, C. M. (1997). Developmental changes in laryngeal and respiratory function with variations in sound pressure level. *Journal of Speech, Language, and Hearing Research, 40,* 595–614.

Stein, L. (1937). *Sprach- und Stimmstörungen und ihre Behandlung in der täglichen Praxis* [*Speech and voice disorders and their treatment in daily clinical practice*]. Vienna-Leipzin-Bern: Weidmann & Co.

Stemple, J. C. (1993). *Voice therapy: Clinical studies*. St. Louis, MO: Mosby Year Book.

Stemple, J. C., Lee, L., D'Amico, B., & Pickup, B. (1994). Efficacy of vocal function exercises as a method of improving voice production. *Journal of Voice, 8,* 271–278.

Story, B. (1995). *Physiologically based speech simulation using an enhanced wave-reflection model of the vocal tract*. Doctoral dissertation, University of Iowa.

Story, B., & Titze, I. R. (1995). Voice simulation with a body-cover model of the vocal folds. *Journal of the Acoustical Society of America, 97,* 1249–1260.

Story, B. H., Titze, I. R., & Hoffmann, E. A. (1996). Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America, 100,* 537–554.

Story, B., Laukkanen, A-M., & Titze, I. R. (2000). Acoustic impedance of an artificially lengthened and constricted vocal tract. *Journal of Voice, 14,* 445–469.

Sundberg, J. (1974). Articulatory interpretation of the "singing formant". *Journal of the Acoustical Society of America, 55,* 838–844.

Sundberg, J. (1987). *The science of the singing voice*. Dekalb: Northern Illinois University Press.

Tapani, M. (1992). *Resonaattoriputki toiminnallisen ääihäirion hoitmenetelmänä. Seitsemän naispotilaan seurantatutukimus* [Resonance tube as a therapy method for a functional voice disorder. A follow-up study of seven female patients] (in Finnish). Helsinki, Finland: University of Helsinki.

Titze, I. R. (1988). A framework for the study of vocal registers. *Journal of Voice, 2*(3), 183–194.

Titze, I. R. (2001). Acoustic interpretation of resonant voice. *Journal of Voice, 15,* 519–528.

Titze, I. R. (2002a). How to use the flow resistant straws. *Journal of Singing, 58,* 429–430.

Titze, I. R. (2002b). Regulating glottal airflow in phonation: Application of the maximum power transfer theorem. *Journal of the Acoustical Society of America, 111,* 367–376.

Titze, I. R. (2004a). A theoretical study of F_0–F_1 interaction with application to resonant speaking and singing voice. *Journal of Voice, 18,* 292–298.

Titze, I. R. (2004b). Theory of glottal airflow and source-filter interaction in speaking and singing. *Acta Acustica-Acustica, 90,* 641–648.

Titze, I. R., Finnegan, E., Laukkanen, A-M., & Jaiswal, S. (2002). Raising lung pressure and pitch in vocal warm-ups: The use of flow-resistant straws. *Journal of Singing, 58,* 329–338.

Titze, I. R., & Story, B. H. (1997). Acoustic interactions of the voice source with the lower vocal tract. *Journal of the Acoustical Society of America, 101,* 2234–2243.

Titze, I. R., & Story, B. H. (2002). Rules for controlling low-dimensional vocal fold models with muscle activities. *Journal of the Acoustical Society of America, 112,* 1064–1076.

Verdolini, K. (2000). Resonant voice therapy. In J. C. Stemple (Ed.), *Voice therapy: Clinical studies* (pp. 46–61). San Diego, CA: Singular.

Verdolini, K., Druker, D. G., Palmer, P. M., & Samawi, H. (1998). Laryngeal adduction in resonant voice. *Journal of Voice, 12,* 315–327.

Verdolini-Marston, K., Burke, M., Lessac, A., Glaze, L., & Caldwell, E. (1995). Preliminary study of two methods of treatment for laryngeal nodules. *Journal of Voice, 9,* 74–85.

Westerman, G. J. (1990). On humming. *Journal of Singing, 46,* 34.

Westerman, G. J. (1996). What humming can do for you. *Journal of Singing, 52,* 37–38.

Received April 27, 2005

Accepted August 22, 2005

DOI: 10.1044/1092-4388(2006/035)

Contact author: Ingo R. Titze, 1101 13th Street, Denver, CO 80204-5319. E-mail: ititze@dcpa.org