Paper presented at THE SECOND INTERNATIONAL CONFERENCE on the PHYSIOLOGY AND ACOUSTICS OF SINGING, October 6-9, 2004, Denver, Colorado. A discussion of each of the figures presented in the oral paper.

The Control of Airflow during Singing

Martin Rothenberg
Professor Emeritus, Syracuse University and President, Glottal Enterprises

Note: The references below that are authored all or in part by the present author can be found on the website www.rothenberg.org.

Figure 1. Typical subglottal air pressures in speech and singing

In singing, subglottal (lung or tracheal) air pressures as high as 40 or 50 cm H_2O have been reported in singing at high volume levels. Such pressures are approximately 3 to 4 times the pressures used for speech. It might be hypothesized that, if not controlled by appropriate compensatory mechanisms, the high airflows that could be caused by such pressures might be detrimental to the mucosa of the vocal folds and would also deflate the lung volume more quickly between breath pauses.

Figure 2. Idealized representations of three ways for reducing glottal airflow during sung **vowels**.

Previous publications have described at least the three mechanisms for reducing glottal airflow in the presence of a high subglottal pressure that are illustrated in this figure. For each method, a hypothetical airflow trace that roughly follows the variation of projected glottal area is shown, in order to provide a comparison that shows what the effect of each mechanism on the airflow waveform is like.

In the first mechanism, sometimes referred to as "pressed" voice, the vocal folds are adducted more than required to sustain voicing. The open quotient and peak airflow would then both decrease.

In the second mechanism, an augmented inertive component of the supraglottal vocal tract impedance suppresses and delays the buildup of airflow during the open phase of the glottal cycle. Since the closing of the glottis at the termination of the glottal open phase forces the flow to zero, there is a net reduction in airflow. Because of the more abrupt termination of airflow caused by this inertive loading of the glottal source, there is also a stronger excitation of the higher glottal harmonics at the instant of vocal fold closure. A good vocal fold closure is required for this mechanism to be effective. (The effect on breathy voice is discussed in reference 3 below.)

In the third mechanism, presumably used by sopranos in the higher pitch ranges, airflow can be suppressed by tuning the vocal tract first or lowest formant (F1) to a frequency at or near the voice fundamental frequency. If the vocal fold closure is good (the voice is not breathy), the open quotient sufficiently small (voice not in falsetto), and the vocal tract resonance sufficiently sharp (production not nasalized, for example), the peaks of supraglottal pressure immediately above the glottis caused by the F1 resonance will occur during the glottal open phase, to oppose the subglottal pressure and thus suppress the glottal flow. (This may be a reason that sopranos do not tune F1 to

higher harmonics, as do male singers, at least at high volume levels. [See the comment of John Nix in the proceedings of this conference.] Tuning to the second harmonic, for example, though increasing the radiated energy at that harmonic, would not be expected to reduce the airflow.)

References:
1. M. Rothenberg, A new inverse-filtering technique for deriving the glottal volume velocity waveform during voicing, J. Acoustical Soc. Amer. 53, 1632-1645 (1973).
2. M. Rothenberg, Acoustic interaction between the glottal source and the vocal tract, in Vocal Fold Physiology, K.N. Stevens and M. Hirano, eds., University of Tokyo Press, 305-328 (1980).
3. M. Rothenberg, Source-tract acoustic interaction in breathy voice, in Vocal Fold Physiology: Laryngeal Function in Phonation and Respiration, T. Baer, C. Sasaki, and K.S. Harris, eds., College Hill Press, San Diego, 254-263 (1984).
4. M. Rothenberg, Cosi' fan tutte and what it means – or – nonlinear source-tract interaction in the soprano voice and some implications for the definition of vocal efficiency, in Vocal Fold Physiology-Laryngeal Function in Phonation and Respiration, T. Baer, C. Sasaki, and K.S. Harris, eds., College Hill Press, San Diego, 254-263 (1986).

Figure 3. Glottal airflow during three types of unvoiced intervocalic consonants.

Illustrated diagrammatically in Fig. 3 are the general patterns to be found for glottal airflow during the predominant types of intervocalic unvoiced consonants. The three classes of consonants illustrated are glottal fricatives (shown as /h/), non-glottal fricatives (shown as /s/), and stop consonants (shown as /t/). During consonants not having a strong vocal tract constriction (as /h/), the airflow would be expected to roughly follow the variation in the glottal area in the abductory movement, though at high degrees of vocal fold abduction the airflow is most often reduced somewhat by the back pressure created by turbulence at the glottis or in the supraglottal vocal tract. In a non-glottal fricative (as in the /s/), the airflow is reduced by the constriction at the point of articulation of the consonant. In a stop consonant (the /t/, shown here aspirated), the airflow is set to zero by the articulatory occlusion for the stop. Since the /t/ is shown aspirated, there is a release of airflow, marked A, after the instant of release that is the aspiration.

Among the ways that airflow can be reduced with a high subglottal pressure during the preceding and following vowels, the possibility exists theoretically that the subglottal pressure can be reduced momentarily during the consonant. We explore that possibility next.

Reference:
1. M. Rothenberg, The glottal volume velocity waveform during loose and tight glottal adjustments, Proceedings of the VII International Congress of Phonetic Sciences, 380-388 (1971).

Figure 4. Subglottal pressure traces illustrating the maximum speed of change for unidirectional and cyclic volitional respiratory gestures.

These traces were collected for reference 1 in order to estimate the dynamic restrictions on changes of subglottal (lung or tracheal) air pressure. They indicate that a cyclic change requires about 300

ms, while a unidirectional change, as an increase or decrease, requires at least about 150 ms if it is to be smooth and not oscillatory.

Reference:
      1. M. Rothenberg, <u>The Breath-Stream Dynamics of Simple-Released-Plosive Production</u>, Bibliotheca Phonetica Vol. 6, Karger, Basel (1968).

Figure 5. Four potentially valid mechanisms for reducing airflow during the production of unvoiced consonants.

Reducing airflow by using the respiratory muscles to reduce the subglottal pressure has been eliminated from the list because of the inordinate dynamic limitations illustrated in Figure 4. The remaining, potentially valid, possibilities are listed in the figure.

Figure 6. Measurements of the variation of airflow during an intervocalic unvoiced consonant with no articulatory obstruction.

This set of airflow traces was recorded to determine the maximum speed at which the vocal folds could be abducted then adducted. In the reference cited, this was referred to as a *cyclic glottal opening gesture*. A gesture of this type is normally a part of the production mechanism of an unvoiced consonant produced intervocalically. The waveforms are of oral airflow during the cyclic abductory gesture of four intervocalic unvoiced consonants, as recorded by a CV mask. The traces during the /p/ consonants (traces A and B) were obtained by bypassing the lip closure with a short length of tubing, to show the airflow that would attain if there were no articulatory closure. Traces C and D were of an intervocalic /h/. Thus all traces at least roughly reflect the changing state of the glottis during the abductory gesture. They were selected from a larger number of productions as typical for the adult male speaker tested.

The traces illustrate that the minimum duration of a cyclic abductory gesture for this speaker (not a trained singer) was about 125 ms if the abduction was to reach the point at which voicing essentially stopped, and about 100 ms if abduction was to result only in a breathy voice (vocal folds vibrating with no vocal fold contact).

Reference:
      1. M. Rothenberg, The Glottal Volume velocity Waveform During Loose and Tight Voiced Glottal Adjustments, <u>Proceedings of the Seventh International Congress of Phonetic Sciences</u>, Mouton, The Hague (1972).

Figure 7. Some potential intervocalic timing patterns for unvoiced stops.

These patterns are extracted from a list in the reference below, which attempts to define a physiologically based phonetic model describing all the phonemically distinct simple-released-plosives that are producible by the vocal tract. In this model it is hypothesized that the categories are determined from the dynamic limitations in rapid speech and discriminability considerations.

Reference:

1. M. Rothenberg, <u>The Breath-Stream Dynamics of Simple-Released-Plosive Production</u>, Bibliotheca Phonetica Vol. 6, Karger, Basel (1968).

Figure 8. Oral airflow patterns comparing an aspirated and unaspirated English /t/ in speech.

The oral airflow patterns in the figure are of an *aspirated* released stop in English (above) and the *unaspirated* released stop (below) produced when joining two English stop phonemes (to produce a single 'geminated' stop articulation).

The traces were recorded with a Glottal Enterprises CV mask system, including the new MS-110 electronics and software. Some low pass filtering (available in the software) is used to clarify the traces by reducing acoustic energy.

Figure 9. The upper trace in Figure 8 contrasted with two examples of the aspirated and unaspirated geminated stop consonants in the spoken phrases "What time" and "What dime".

The traces in the figure were positioned so as to align the instants of articulatory release at the vertical line.

The geminated sequence /tt/ across a word boundary produces an aspirated release similar to that of a single aspirated /t/. The sequence /td/ across a word boundary showed an articulatory release closely synchronized with the adduction of the vocal folds for the succeeding vowel, to produce an unaspirated release, and thus reduce the volume of air expended. The closure in both instances was held longer than for the single consonant, presumably in order to signal the presence of two consonants in the underlying phoneme sequence.

Figure 10. Three airflow traces in which a phoneme /s/ is followed by an unvoiced consonant /p/.

The traces are recorded with the same Glottal Enterprises mask system as used in Figures 8 and 9. The traces show how the duration of the vocal fold abductory movement is used to signal juncture and control aspiration. In the non-word-initial /p/ of the production of the English word "spot", there is no aspiration, as dictated by English phonology; the glottal adduction is closely synchronized to the articulatory release. The lower trace shows how this characteristic of English can be used by an English speaker who has not mastered the purposeful control of aspiration in singing (or the pronunciation of a foreign language) to reduce the aspiration in the release of the /p/ in the phrase "This pot". By visualizing the phrase "Thi spot" (similar to the real phrase "The spot"), a potentially perceptually acceptable pronunciation of "This pot" is produced with the /p/ not aspirated, and expired airflow greatly reduced.

Figure 11. Comparison of airflow in spoken and sung unvoiced consonants. (Adapted from the reference below.)

Figure 11 shows a comparison of CV mask airflow traces from the same nonsense sentence, when spoken at a moderate volume level and when sung loudly near the top of the range of this singer. The pressure trace at the top of the figure was recorded from a slightly inflated balloon placed in the

esophagus via a small diameter flexible tube introduced at the nares. Such a procedure can yield a rough representation of the subglottal (tracheal) pressure if there are no esophageal contractions present. The airflow trace was low-pass filtered to remove the blur that would be caused by voicing. Both flow and esophageal pressure traces during singing show vibrato-related oscillations, which can be neglected in the analysis.

This singer appears to have employed some of the mechanisms discussed above to reduce the expended air during the unvoiced consonants. For example, the duration and extent of the abductory gesture for both instances of /h/ appear to have been reduced to the extent that the air expended in each was less than in the spoken versions. Also, the word-initial /p/ in "pat", while correctly aspirated in speech, was only slightly aspirated in the sung version.

Reference:
1. M. Rothenberg, D. Miller, R. Molitor and D. Leffingwell, The Control of Airflow in Loud Soprano Singing, J. of Voice, Vol. 1, No. 3, 262-268 (1987).

****