Estimating perceived phonatory pressedness in singing from flow glottograms

Sundberg, J. and Thalén, M. and Alku, P. and Vilkman, Erkki

journal: TMH-QPSR
volume: 43
number: 1
year: 2002
pages: 089-096

http://www.speech.kth.se/qpsr
Estimating perceived phonatory pressedness in singing from flow glottograms

Johan Sundberg*, Margareta Thalén**, Paavo Alku***, Erkki Vilkman****

* Speech Music Hearing department, KTH, Stockholm
** SMI (University College of Music Education in Stockholm)
*** Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Box 3000, Fin-02015 TKK, Finland
**** University of Oulu, Department of Otolaryngology and Phoniatrics, Oulu, Finland, and Helsinki University Central Hospital, Box 220, Fin-00029 HUS, Finland

Paper presented at the 31st Annual Symposium Care of the Professional Voice, Philadelphia, June 2002

Abstract
The normalized amplitude quotient (NAQ), defined as the ratio between the peak-to-peak amplitude of the flow pulse and the negative peak amplitude of the differentiated flow glottogram and normalized with respect to period time, has been shown to be related to glottal adduction. Glottal adduction, in turn, affects mode of phonation and hence perceived phonatory pressedness. The relationship between NAQ and perceived phonatory pressedness was analyzed in a material collected from a professional female singer and singing teacher who sang a triad pattern in breathy, flow, neutral and pressed phonation in three different loudness conditions (soft, middle, loud). In addition, she also sang the same triad pattern in four different styles of singing, classical, pop, jazz and blues, in the same three loudness conditions. A panel of nine experts rated the degree of perceived pressedness ratings with the mean NAQ values for the various triads showed that about 73% of the variation in perceived pressedness could be accounted for by variations of NAQ.

Introduction
The voice source can be continuously varied with regard to vocal loudness and pitch, related to the physiological voice control parameters subglottal pressure (P_s) and vocal fold length and stiffness, respectively. In addition, the voice source is affected by the degree of glottal adduction; a firm glottal adduction produces a pressed/hyperfunctional voice and a weak glottal adduction gives a breathy/hypofunctional voice. Glottal adduction is relevant to vocal hygiene; a habitually exaggerated glottal adduction is known to frequently lead to voice disorders such as nodules.

A change of the glottal adduction affects the voice source in various ways. Increased adduction tends to decrease the vibration amplitude of the vocal folds. This reduces the amplitudes of the flow pulses and increases the closed quotient. More specifically, it changes the relationship between Ps and various characteristics of the voice source waveform. This relationship, in turn, is reflected in various measures. For example, the estimated projected glottal area, defined as the ratio between the peak-to-peak pulse amplitude and the square root of Ps, decreases, when glottal adduction is increased. Glottal compliance, i.e., the ratio between the air volume displaced in a pulse and the associated Ps also decreases with increasing glottal adduction. An increased glottal adduction also reduces the level difference between the two lowest partials of the voice source spectrum. Moreover, the amplitude quotient (AQ), referred to as Td by Fant (1997), has been shown to yield a measure that faithfully reflects the degree of glottal adduction (Alko & Vilkman, 1996). AQ
is defined as the ratio between the peak-to-peak flow pulse amplitude and the negative peak amplitude of the differentiated flow glottogram. Since AQ yields a time-domain measure that can be interpreted as a length of the sub-section of the glottal closing phase, it is justified to normalize this quotient by dividing it with the length of the fundamental period (Alku et al., 2002). Similar parameterization scheme has also been used in Fant (1997). The normalized version of AQ is denoted by the Normalized Amplitude Quotient (NAQ).

Glottal adduction differs between singing styles. While in the classical operatic tradition, adduction seems to be reduced to a minimum, it is often quite firm in certain styles of singing practiced in popular music, e.g. country (Sundberg et al., 1999b), Broadway (Stone et al., to be printed) and belting (Sundberg et al., 1993; Bestebreurtje & Schutte, 2000). In a recent investigation, co-authors JS and MT analyzed the voice source in different styles of singing as produced by a professional female singer (Sundberg & Thalén, 2001). The singer subject also produced a set of examples where she deliberately attempted to vary the degree of glottal adduction. This material was then submitted to a listening test where expert subjects rated the degree of phonatory pressedness. In the present investigation, we will compare these ratings with different voice source measures of glottal adduction.

Material

The subject (co-author MT) has a long professional experience of singing and teaching in the four different styles: Classical, Pop, Jazz and Blues. Originally trained in the Classical style of singing she performed professionally as a soloist, mostly in the Jazz, Pop and Blues styles for about 10 years. In addition, she has worked as a voice teacher in these styles for more than 20 years.

The subject sang, in one breath group, the syllable [pæ] twice on each of the pitches of a seventh chord (A3, C#4, E4, G4). She sang this pattern in four modes of phonation: breathy, flow, neutral, and pressed, each in three degrees of vocal loudness (soft, middle and loud). (Flow phonation is produced with the lowest degree of glottal adduction that produces a vocal fold closure and is typically used in Classical singing, apparently corresponding to what has been referred to as “resonant voice”. (Verdolini et al., 1998)). The subject also sang the same material in four styles: Classical, Pop, Jazz, and Blues, again each in three degrees of vocal loudness (soft, middle and loud). The Classical style of singing was that typically used in performances of the Lieder repertoire. The Pop and Jazz styles corresponded to those typically used in Pop and Jazz ballads. The Pop style is typically represented by performers like Randy Crawford and Whitney Houston. Typical representatives of the Jazz and Blues styles are performers like Billie Holiday and Sarah Vaughan, and Bessie Smith and Janis Joplin, respectively. The subject sang the pitch pattern in the following order: breathy, flow, neutral, pressed, Classical, Pop, Jazz, Blues, starting with low, then medium and last high degree of vocal loudness.

The flow signal was obtained from a flow mask (Glottal Enterprises, MSIF2). The audio signal was recorded by a high fidelity microphone at a distance of 0.3 m from the mouth. Ps was recorded as the oral pressure during the occlusion of the consonant /p/ in the syllable [pæ:]. This pressure was measured with a pressure transducer connected to a small plastic tube, mounted in the flow mask, that the subject had in the corner of her mouth. All these signals were recorded on separate channels in a TEAC PCM recorder.

Perceptual analysis

The recorded material was evaluated perceptually by a panel of 10 listeners, all experts on voice pedagogy and acquainted with different singing styles. The task was to assess, on 100-mm long visual analogue scales (VAS), the degree of phonatory pressedness in all styles and modes of phonation. The ends of the VAS were marked “Extremely breathy” and “Extremely pressed”. These terms were familiar to the listeners.

Each of the panel members received a copy of an audiotape with the stimuli in the same randomized order. In the tape, each of the triads occurred two times (4 phonation modes x 3 loudnesses x 2 presentations plus 4 styles x 3 loudnesses x 2 presentations).
Table 1 lists the correlation between the ratings of the first and second presentation obtained for identical stimuli from each rater. Also shown are the slope and the intercept of the best linear fit of the data points. The correlation averaged across raters for the responses to repeated stimuli was 0.794 (SD=0.295). Subject #8 yielded an exceptionally low correlation. The responses of this subject were therefore discarded.

Table 1. Intra-rater consistency for ratings of phonatory press according to the Pearson correlation coefficient r. Slope and Itcpt show the slope and intercept of the best linear fit to the data.

<table>
<thead>
<tr>
<th>Subject #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope</td>
<td>0.98</td>
<td>1.13</td>
<td>1.00</td>
<td>0.62</td>
<td>0.51</td>
<td>0.91</td>
<td>1.00</td>
<td>0.26</td>
<td>0.98</td>
<td>0.92</td>
</tr>
<tr>
<td>Itcpt</td>
<td>2.3</td>
<td>-11.9</td>
<td>-4.2</td>
<td>22.0</td>
<td>22.8</td>
<td>2.6</td>
<td>-4.3</td>
<td>40.6</td>
<td>-4.4</td>
<td>2.68</td>
</tr>
<tr>
<td>r</td>
<td>0.974</td>
<td>0.969</td>
<td>0.932</td>
<td>0.669</td>
<td>0.562</td>
<td>0.877</td>
<td>0.961</td>
<td>0.216</td>
<td>0.855</td>
<td>0.926</td>
</tr>
</tbody>
</table>

Table 1 lists the correlation between the ratings of the first and second presentation obtained for identical stimuli from each rater. Also shown are the slope and the intercept of the best linear fit of the data points. The correlation averaged across raters for the responses to repeated stimuli was 0.794 (SD=0.295). Subject #8 yielded an exceptionally low correlation. The responses of this subject were therefore discarded.

Figure 1 shows the panel’s mean rated pressedness averaged across the pitches contained in each triad. For both modes and styles, the mean rated pressedness was lowest for the soft triads and highest for loud triads except for the pressed mode and the Blues style, where the differences were small and showed a weak opposite trend. For the modes, the results showed the lowest degree of rated pressedness for breathy and the highest for pressed, as expected. For the styles, Classical showed the lowest and Blues the highest degree of pressedness.

**Acoustic analysis**

A block-scheme of the analysis equipment is shown in Figure 2. The voice source waveform was derived by inverse filtering the flow signal (Glottal Enterprises, MSIF2). The lowest pitch in each seventh chord was captured on a transient recorder (Glottal Enterprises, BT-1) and the frequencies and bandwidths of the two lowest formants were adjusted so as to minimize
the ripple during the closed phase. Then, all pitches of the seventh chord were run through the inverse filter and stored in a file, using the Soundswell signal workstation (Ternström, 1992). Ps was simultaneously recorded on another channel of the same file.

For each note in the triad a representative waveform was selected from the middle part of the second syllable, Figure 3. The glottogram characteristics as well as Ps were then measured using the Soundswell signal workstation. As illustrated in Figure 4, the following flow glottogram parameters were measured:

- period time between two adjacent discontinuities corresponding to the closing of the glottis,
- closed phase between the end of a flow pulse to the onset of the next flow pulse,
- glottal leakage, defined as the mean flow amplitude during the closed phase,
- peak-to-peak pulse amplitude, defined as the peak value of the pulse minus the glottal leakage, and

![Figure 2. Block diagram of the equipment used during the analysis.](image)

![Figure 3. Example of sound file for the pitches C#4 and A3 sung in Blues style. The upper and lower traces show oral pressure and flow, respectively.](image)
air volume contained in the pulse, estimated from a triangle approximation of the pulse.

Ps was determined from the [p] preceding the syllable analyzed; as the subject kept vocal loudness constant during the syllable analyzed this value of Ps was assumed to be representative of the entire subsequent syllable.

The glottogram files were differentiated using the Extract module of the Swell package, Figure 4b. The level difference between the first and second voice source spectrum partials, henceforth H1-H2, was measured by means of a narrow bandwidth FFT analysis of the flow glottogram analyzed.

From these data the following parameters were computed:

1. relative estimated glottal area defined as the ratio between the peak-to-peak pulse amplitude and the square root of Ps,
2. closed quotient,
3. glottal compliance, defined as the ratio between the air volume contained in the glottal pulse and Ps, and
4. normalized amplitude quotient, defined as the ratio between peak-to-peak pulse amplitude and the negative peak amplitude of the differentiated flow glottogram divided by the length of the fundamental period.

Thus, one physiological measure was studied, Ps, and five measures related to glottal adduction, relative estimated glottal area, H1-H2, closed quotient (QClosed), glottal compliance, and normalized amplitude quotient (NAQ).

Results

Figure 5 shows the relationships between the five adduction measures and the mean rated degree of phonatory pressedness. Of these, the relative estimated glottal area showed the lowest correlation. Among the other measures, H1-H2
showed the lowest ($R^2 = 0.660$) while Qclosed, glottal compliance and NAQ showed the highest ($R^2 = 0.718$, $R^2 = 0.755$, and $R^2 = 0.725$, respectively). These results suggest that most of the variation in perceived phonatory pressedness can be accounted for in terms of the latter three measures. Of these, NAQ seems particularly interesting, being a robust measure that can be derived automatically from high fidelity audio recordings (Alku et al., 2002). In addition, it is closely related to perceived pressedness, accounting for 73% of the variation in perceived pressedness.

According to these results, the mean NAQ can be regarded as a reasonable approximation of the degree of phonatory pressedness. It is then interesting to specify the phonatory characteristics of the styles and phonation modes in terms of mean degree of pressedness and mean loudness. This can be realized in terms of a phonation map, showing mean NAQ versus mean $P_s$. Figure 6 shows the results in terms of such a phonation map. Breathy and pressed were located at opposite extremes, both with respect to pressedness and $P_s$. Neutral and flow share mean $P_s$ with breathy, but the mean NAQ is lower in neutral. With regard to the styles, Classical is close to breathy, and Blues close to pressed, although the latter had a lower NAQ value. Pop and Jazz have similar mean $P_s$ but

Figure 5. Relationship between the five adduction measures and the mean rated degree of phonatory pressedness. The best linear fit of the data points (lines) are shown by the equations and the associated $R^2$ values for the correlation.
Jazz has a considerably lower mean NAQ, midway between neutral and pressed, while the NAQ for Pop was close to that of flow phonation.

**Discussion**

Normalized amplitude quotient is closely related to other flow glottogram measures. For example, the peak-to-peak pulse amplitude has been found to be strongly correlated with the amplitude of the voice source fundamental (Gauffin & Sundberg, 1989; Fant, 1997). Likewise, Ps has been found to strongly correlate with the negative peak amplitude of the differentiated flow glottogram (Sundberg et al., 1999a). This implies that basically NAQ should reflect the relationship between the amplitude of the fundamental and Ps. On the other hand, those relationships have been documented only for a limited variation of phonatory modes in previous investigations. Therefore, the present material would represent a more valid test of the relationship between perceived degree of pressedness and the various glottogram characteristics.

The phonation map is an attempt to describe, in a compact form, main phonatory characteristics of modes of phonation and singing styles. It has the advantage of taking into account two main physiological variables for phonatory control, Ps and the degree of glottal adduction. Co-authors JS and MT recently presented a similar map, where, however, the vertical axis represented a pressedness factor, derived as a weighted sum of Qclosed, glottal compliance and H1-H2. This measure showed a correlation of R²=0.876 with the mean pressedness ratings, while NAQ correlation was R²=0.725. On the other hand, the weights for the three terms in the pressedness factor were optimized with respect to the correlation with the pressedness ratings. Therefore, we believe that the NAQ measure is more useful than the pressedness factor.

The locations of the phonation modes and the styles of singing in the phonation map appeared, by and large, logical. Blues was found to be close to pressed, however slightly less pressed. This difference may be important from the point of view of vocal health. The small difference between breathy and Classical seems more problematic, since breathy phonation must be produced with less adduction than Classical.

The present investigation was based on data from one singer subject. As the results are promising, it would be worthwhile to apply a similar research strategy on some more voices.

**Conclusions**

The normalised amplitude quotient, available by automatic analysis of the audio signal, shows a high correlation with perceived degree of phona-
tory pressedness. It seems useful for a compact description of phonatory characteristics of different modes of phonation and styles of singing. Such description can be given the form of a phonation map, showing mean normalised amplitude quotient versus mean $P_r$.

Acknowledgement

Co-author MT’s participation in this investigation was supported by a grant within the Research and Development in the Arts program at the SMI (University College of Music Education in Stockholm). The inverse filtering and the glottogram measurements were made by Jenny Iwarsson.

References


