ESTIMATION OF OUTPUT-COST–RATIO USING AN AEROELASTIC MODEL OF VOICE PRODUCTION

J. Horáček¹, A-M. Laukkanen², P. Šidlof³

¹Institute of Thermomechanics, Academy of Sciences of the Czech Republic, Prague, Czech Republic
²Department of Speech Communication and Voice Research, University of Tampere, Tampere, Finland

Abstract: The study deals with mathematical modeling of the vocal fold self-oscillations related to estimation of the so-called output-cost-ratio (OCR), which is computed from the numerically simulated sound pressure level at the glottal level and the impact stress (IS) during vocal folds collision. The dependence of OCR on prephonatory glottal width, fundamental frequency and lung pressure is discussed and partly compared with a modified output cost ratio measured in humans, where the closed quotient is used instead of IS.

Key words: Biomechanics of voice, numerical simulation of vocal folds vibration.

I. INTRODUCTION

Impact stress (IS, i.e. the impact force divided by the contact area) has been regarded as the main loading factor in voice production and the most plausible cause of vocal fold traumas like nodules. To quantify the cost of voice production, Berry et al. [1,2] presented a parameter called output-cost-ratio (OCR), which concerns the acoustic output in relation to IS:

\[ OCR = 20 \log \frac{P_{\text{sup}}}{P_0} - 20 \log \frac{\text{IS}}{\text{IS}_0}, \]

where \( P_{\text{sup}} \) is supraglottal acoustic sound pressure (measured at a distance of 15 cm above the glottis of an excised canine larynx), \( P_0 \) and \( \text{IS}_0 \) are constants.

IS is difficult to measure directly in humans. The present study investigates the output-cost-ratio using an aeroelastic model of voice production. The aeroelastic model of vocal folds vibration enabled to study the output-cost-ratio OCR in more details than in the experiments with the excised larynges. The influence of various parameters on OCR can be studied separately and in a more controllable way.

It has been found that closed quotient (CQ, i.e. closed time of the glottis divided by the period length) obtained from electroglottographic (EGG) signal correlates with IS - see Verdolini et al.[9]. Laukkanen et al. [7] have tested in human subjects the so-called Quasi-Output-Cost ratio where CQ has been used instead of IS.

The present study compares results of OCR obtained with modelling to some of the results obtained for human subjects by Laukkanen et al [7].

II. METHOD

IS magnitudes and sound pressure level (SPL\text{source}) above the glottis were quantified using an aeroelastic computer model of the vocal fold self-oscillations employing the Hertz model of impact forces during vocal fold collision - see [4,5]. The model is based on a two-degrees-of-freedom dynamic system allowing rotation and translation of the vocal-fold-shaped element vibrating on two springs and dampers - see Fig. 1. Self-oscillations are excited by nonlinear aerodynamic forces resulting from the fluid-structure interaction.

The impact Hertz force is given as 

\[ F_H = k_H \delta^{3/2}, \]

where \( k_H \) is the contact stiffness and \( \delta \) is the penetration of the vocal fold through the symmetry axis during collision. IS was calculated as the maximum value during one oscillation period according to the formula:

\[ IS = \frac{3}{2} \frac{F_{H,\text{max}}}{\pi a^2}, \quad a = \frac{3}{4} r \frac{(1-\nu^2)}{E} F_{H,\text{max}}, \]

where \( F_{H,\text{max}} = k_H \delta_{\text{max}}^{3/2}, \quad k_H = \frac{4}{3} \sqrt{r \frac{E}{1-\nu^2}}, \quad r \)

is the radius of the curvature of the vocal fold model at the contact point, \( E \) is Young modulus and \( \nu \) is Poisson number; for \( E = 8000 \text{ Pa}, \ \nu = 0.4 \).

A parabolic shape of the vocal fold surface was considered, which gives the radius \( r \). For the on-line numerical simulations in time domain, the resulting system of four 1st order ordinary differential equations describing the vocal fold vibrations was solved by the 4th order Runge-Kutta method.

**Figure 1.** Schema of the aeroelastic model – [4]

In calculating OCR with the model, values of \( P_0 = 20 \mu\text{Pa} \) and \( \text{IS}_0 = 1 \text{ Pa} \) were used. Prephonatory glottal...
half-width was set as $g=0.2-0.5$ mm, i.e. the glottal width varied between 0.4 and 1 mm. Fundamental frequency $F0$ was set to 100 and 400 Hz. Using the model presented here, it was not possible to use negative pre-phonatory glottal width (corresponding to pressed phonation), which Berry et al. [1,2] also used. In the present study, the lung pressure ($P_{lung}$) and airflow values were set within the range reported for healthy humans, ($P_{lung}\leq3000$ Pa, airflow rate $Q<0.8$ l/s – see Hirano [3]). The computations were realized in the range of $P_{lung}$ from the phonation threshold pressures ($P_{th}$) to the phonation instability pressure ($PIP$).

In measurements, the data were obtained from human subjects (see - [7]). The subjects were 62 females producing [pa:p:a] 5 times loudly. The sound pressure level ($SPL$) was registered at 40 cm from the subject’s lips, closed quotient $CQ_{EGG}$ was calculated from EGG signal. The acoustic signal was recorded using a digital recorder and B&K 4164 microphone, and EGG signal was registered with Glottal Enterprises dual-channel EGG. Oral pressure was registered with MSIFT-II (Glottal Enterprises). The oral pressure during voiceless plosive [p] was used as an estimate of subglottic pressure. The prephonatory glottal half-widths $g$ were set within the range reported for healthy humans, ($P_{lung}\leq3000$ Pa, airflow rate $Q<0.8$ l/s – see Hirano [3]). The subjects were 62 females producing [pa:p:a] 5 times loudly. The sound pressure level ($SPL$) was registered at 40 cm from the subject’s lips, closed quotient $CQ_{EGG}$ was calculated from EGG signal. The acoustic signal was recorded using a digital recorder and B&K 4164 microphone, and EGG signal was registered with Glottal Enterprises dual-channel EGG. Oral pressure was registered with MSIFT-II (Glottal Enterprises). The oral pressure during voiceless plosive [p] was used as an estimate of subglottic pressure. The acoustic signal was analyzed for mean $F0$ and $SPL$ using Intelligent Speech Analyser (ISA) signal analysis device (developed by Raimo Toivonen, M.Sc. Eng). $CQ_{EGG}$, vibration period $T$ ($F0=1/T$) and the mean oral pressure during [p] were measured by using a custom-made program for measurement of AC- and DC signals (developed by Heikki Alatalo, DSP-Systems).

### III. RESULTS AND DISCUSSION

Figure 2 shows the simulated $SPL_{source}$ values, at the upper end ($x=L$) of the glottis, for all considered prephonatory glottal half-widths $g$ as a function of lung pressure, which is presented as a dimensionless normalized excess subglottal pressure

$$P_{sen} = \frac{P_{lung} - P_{th}}{P_{th}}$$  \hspace{1cm} (3)

As expected, after crossing the phonation onset at the phonation threshold pressure $P_{th}$, the $SPL_{source}$ increases with the pressure $P_{sen}$ for all $g$ values in a nearly linear way. The highest $SPL_{source}$ values are reached for $g=0.5$ mm near the $PIP$, where the lung pressure values are at a maximum.

The $IS$ values obtained with the model (see Fig. 3) are in the range of the data reported for living subjects and excised human and canine hemilarynges (see - [1,6,9]). $IS$ increased with the lung pressure reaching a plateau when getting close to the $PIP$ values. Again, the maximum values of $IS$ were obtained for $g=0.5$ mm near $PIP$ where also a lung pressure maximum occurs. Nearly zero $IS$ values are near $P_{th}$, i.e. near $P_{sen} = 0$.

The $OCR$ calculated according to the equation (1) from the simulated $SPL_{source}$ and $IS$ values is shown in Fig. 4.

The maximum of $OCR$ appears near $P_{sen} = 0$ due to the very low $IS$ values near the phonation threshold. For all prephonatory glottal half-widths $g$, the $OCR$ decreases with $P_{sen}$, having minimal values at about $P_{sen} \approx 1.5$, thereafter the $OCR$ values slightly increase up to the $PIP$ values, where the $IS$ reaches a plateau, while $SPL_{source}$ still increases (compare Fig. 4 with Figs. 2,3). We can note that according to the model and the definition (1) of the $OCR$ parameter, the most advantageous (economic) regime would be to phonate near the phonation onset. It seems to be a peculiar but trivial and expected result, because at $P_{th}$ there are none or very small impacts ($IS \rightarrow 0$ ) and therefore $OCR$ theoretically goes to the infinity ($OCR \rightarrow +\infty$).

![Figure 2. Computed SPLsource versus normalized excess subglottal pressure $P_{sen}$ (F0=100 Hz).](image2)

![Figure 3. Computed IS versus Psen. (F0=100 Hz).](image3)

![Figure 4. Computed OCR versus Psen. (F0=100 Hz).](image4)
The OCR values varied with the prephonatory glottal width \( g \) in dependence on \( P_{\text{lung}} \) in a qualitatively different way, as the present settings were used (see Fig. 5). According to the results by Berry et al. \[1\] for \( F_0=150 \) Hz, a prephonatory glottal width of 2 mm was optimal (gave the largest SPL with the lowest IS) in excised canine larynges, while with their model of the vocal folds with a vocal tract, a width of 1 mm was optimal (OCR values reached the maximum). Later Berry et al. \[2\] reported a broad maximum in the OCR curves at about 0.6 mm for excised canine larynges when \( P_{\text{sub}} \) was varied in the range 1 – 1.6 kPa. The results of the present study suggest that the optimal glottal width is dependent on the lung pressure (see Figs. 4 and 5). At low \( P_{\text{lung}} \) values a larger prephonatory glottal width seems to be more economic, while at high \( P_{\text{lungs}} \) values a smaller width would be more preferable. It should be noted, however, that using the present aeroelastic model, phonation with really small glottal widths (corresponding to pressed phonation) was not possible to model.

Because IS is difficult to measure in humans, CQ may be used as a substitute for it, based on the fact that there is a correlation between CQ and IS reported in excised canine larynges \[9\]. The relation between IS and CQ obtained with the model of the present study is shown in Fig. 6. In general, we can suppose the relation in the following form:

\[
IS = a \ CQ^b, \tag{4}
\]

where \( a \) and \( b \) are constants dependent on \( g \) as shown in Fig. 6. The exponent varied from \( b=1.2 \) to 3.7 in dependence on the prephonatory glottal half-width.

After substituting IS from equation (4) to the formula (1) the OCR can be approximated by a Modified Output Cost Ratio parameter defined as

\[
\text{MOCR} = \text{SPL}_{\text{source}} - 20 \times 2 \log CQ + \text{const.}, \tag{5}
\]

where the constant \( b \) in equation (4) was approximated by the value \( b=2 \) for all prephonatory glottal half-widths considered. The computed MOCR is shown as function of the normalized subglottal pressure in Fig. 7.

The MOCR parameter calculated from the data measured in humans is shown in Fig. 8 as function of the subglottal (oral) pressure. The trend, i.e. the increase of MOCR with \( P_{\text{sub}} \) is in good agreement with the modeled data presented in Fig. 7 for the higher \( P_{\text{sen}} \) values.

\[
y = 151278 \times 1.28, \quad y = 11898 \times 1.32, \quad y = 9455.3 \times 1.39, \quad y = 374217 \times 3.73
\]

\[
0 \quad 0.1 \quad 0.2 \quad 0.3 \quad 0.4 \quad 0.5
\]

\[
CQ \quad IS \text{ (Pa)}
\]

\[
g=0.2mm \quad g=0.3mm \quad g=0.4mm \quad g=0.5mm
\]

Figure 6. Computed IS versus CQ for various prephonatory glottal half-widths. \((F_0= 100 \) Hz).
of MOCR changes from a maximum at $P_a$ pressure through a minimum to another maximum at $PIP$ pressure values are similar for both $F0$ values, however, the values of MOCR are higher for the higher fundamental frequency $F0=400$ Hz.

![Figure 9](image1.png)

**Figure 9.** Computed MOCR versus $P_{long}$ for various prephonatory glottal half-widths $g$ for $F0=100$ and 400Hz.

![Figure 10](image2.png)

**Figure 10.** Measured MOCR versus fundamental frequency $F0$ in humans - [7].

The modelled influence of the fundamental frequency is in good qualitative agreement with the data measured in humans as shown in Fig.10, where the increase of MOCR parameter with $F0$ is obvious.

IV. CONCLUSIONS

The present study tested an output-cost ratio parameter which is supposed to reflect economy of voice production. An aeroelastic model of vocal fold vibration and material recorded from female subjects were used. Results obtained with modeling corresponded to those obtained from humans.

Based on the results, it looks like that the rate of SPL rise in relation to $P_{sub}$ and $F0$ exceeds the rise in IS. This results in the fact that OCR does not correspond to the clinical and pedagogical observations suggesting that using loud phonation and high pitch (for an excessively long time) increases the risk of vocal fatigue and vocal fold traumas. A more complicated parameter, taking into account the effects of $F0$ (=increased number of collisions in time), loading aerodynamic and inertia forces caused by the acceleration of the vocal fold tissue might better reflect the mechanical vocal fold loading and thus better describe the economy of voice production.

ACKNOWLEDGEMENTS

The research was supported by the Grant Agency of the Academy of Sciences of the Czech Republic (Project No. IAA206401 Mathematical modeling of human vocal fold oscillations), by the Academy of Finland (Grant No. 106139 Biomechanical study on the traumatizing mechanisms in vocal fold vibration), and by COST 2103 Action Advanced Voice Function Assessment.

REFERENCES


