Modeling source-tract interaction in speech production: Voicing onset vs. vowel height after a voiceless obstruent

Jorge C. Lucero¹, Laura L. Koenig², Susanne Fuchs³

¹Dept. Computer Science, University of Brasilia, Brasilia DF 70910-900, Brazil
²Haskins Laboratories, 300 George Street, New Haven CT 06511, USA
³Center for General Linguistics (ZAS), Schuetzenstrasse 18, Berlin 10117, Germany

lucero@unb.br, koenig@haskins.yale.edu, fuchs@zas.gwz-berlin.de

Abstract

This paper shows that observed higher values of intraoral pressure at voicing onset of higher vowels vs. low vowels might be consequence of the acoustical coupling between the vocal fold oscillation and the vocal tract. The acoustical coupling is characterized by a simple model based on a lumped description of tissue mechanics, quasi-steady flow and one dimensional acoustics. The model predicts a “U”-shaped relation between the transglottal pressure at voicing onset and ratio of fundamental frequency to the first formant, which reproduces the data with good approximation.

Index Terms: voicing onset, transglottal pressure, acoustical coupling, vowel height

1. Introduction

A recent study on voicing onset and offset around voiceless obstruents found that voicing onset occurred at higher levels of intraoral pressure ($P_{io}$) before high vowels (/i/ and /u/) than low vowels (/a/) [1].

The cause of this effect is not clear. In fact, it is well known that high vowels are produced with slightly higher fundamental frequency ($f_0$) than low vowels (“intrinsic $f_0$ effect”) [2]. Phonation at higher frequency demands a larger transglottal pressure to start [3], so lower values of $P_{io}$ would be expected (for a fixed subglottal pressure). The recent speech data [1] showed, indeed, the intrinsic $f_0$ effect. Nevertheless, phonation onset $P_{io}$ did not follow the expected pattern.

Higher vowels have also been associated with lower degrees of vocal fold contact [4], lower speed quotients [5] and longer voice onset times [6]. Those effects would also lead one to expect lower $P_{io}$’s.

On the other hand, past modeling work suggests that the acoustical coupling between the vocal fold oscillation and the vocal tract might increase phonation onset $P_{io}$ for high vowels. When $f_0$ is below the first formant $F_1$, the air column in the vocal tract behaves as an inertive load and decreases the phonation threshold subglottal pressure [7]. For a fixed $f_0$, the lower $F_1$ the larger the inertive reactance; therefore, a high vowel configuration (low $F_1$) would demand a lower transglottal pressure to start the vocal fold oscillation than a low vowel (high $F_1$). A similar result has been obtained from a mechanical replica of the vocal folds and vocal tract [8]. The sound wave reflected back at the lips towards the glottal source influences the vocal fold oscillation. Thus, the transglottal pressure to start the oscillation depends on the length of the vocal tract tube ($l$). Starting from a very short tube, an increase of its length causes a decrease of the transglottal pressure, until reaching a minimum when $f_0$ is about half the first resonance (formant) of the vocal tract ($F_1 = c/(4l)$, where $c$ is the speed of sound). The experimental results were theoretically characterized using Titze’s mucosal wave model [7] coupled to a waveguide analogy of the vocal tract [9, 10].

This paper will investigate if the observed $P_{io}$ pattern may be attributed to the acoustical coupling with the vocal tract, using the theoretical model applied to the mechanical replica mentioned above.

2. Data

Intraoral pressure was recorded from eight adult speakers of Standard German, of ages between 25 and 45 years, using a pressure sensor located at the palate [1]. The speakers were three females (W1-W3) and five males (M1-M5) and they produced words with target consonant-vowel sequences embedded in a carrier sentence. The target consonants were /t/, /tʃ/, /ts/, /l/ and /st/, and they were followed by either the low vowel /a/ or the high vowels /i/ and /u/. A total of twelve sentences was used, and each sentence was produced ten times by each speaker. After removing productions with disfluencies and other problems, a final data set of 762 tokens was obtained.

The time of voicing onset after the consonant was determined by visual inspection of the measured $P_{io}$ signal. Values of $f_0$ and $F_1$ after voicing onset were computed using Praat software. Then, the $P_{io}$ signal was smoothed at a 40 Hz cut-off frequency to extract the glottal pulses, and corrected for baseline drift. The smoothed signal was
used to determine the peak value of $P_{io}$ during the obstructed and its value at voicing onset.

Figs. 1 and 2 show statistical results of the measured data for each subject. In each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the circles represent outliers. All subjects show an increased $P_{io}$ for high vowels vs. low vowels (Fig. 1). If the peak $P_{io}$ during the obstructed is regarded as an approximate representation of the subglottal pressure, then subtracting the onset $P_{io}$ from it yields the transglottal pressure ($P_{trans}$) at voicing onset. All subjects show a lower value of onset $P_{trans}$ for high vowels vs. low vowels. Also, all subjects show an intrinsic $f_0$ effect in the expected direction (Fig. 2).

3. Theoretical model

The vocal folds are characterized using Titze’s mucosal wave model [7]. The tissue biomechanics follows the equation of motion

$$M\ddot{x} + B\dot{x} + Kx = P_g,$$

where $x$ is the tissue displacement at the midpoint of the glottal channel, $M$, $B$ and $K$ are the mass, damping and stiffness, respectively, per unit area of the vocal fold medial surface, and $P_g$ is the mean glottal air pressure

$$P_g = P_s + (P_s - P_i)(1 - a_2/a_1)/k_t,$$

where $P_s$ is the subglottal pressure, $P_i$ is the supraglottal pressure, $1 \leq k_1 \leq 1.4$ is a transglottal pressure coefficient, and $a_{1.2} = 2L[x_0 + x(t \pm \tau)]$ are the glottal areas at the lower and upper edges of the glottal channel, respectively, where $x_0$ is the glottal half-width at rest, $\tau$ is the time delay for the mucosal wave to travel half the glottal height, and $L$ is the vocal fold length.

The vocal tract is represented as a hard-walled tube or waveguide of length $L$. The supraglottal pressure $P_i$ (at the epilarynx) is the sum of two components: a static pressure $P_{st}$ and an acoustic pressure $P_i^a$. In our analysis, we will ignore any static pressure drop between the epilarynx and the mouth cavity, and will assume $P_i = P_{io}$.

The acoustic pressure is the resultant of an incident pressure wave injected from the glottis, denoted $p_i^a$, and a backward propagating wave after partial reflection at the other end of the tube (the open mouth), denoted $p_i^r$.

The incident pressure wave is produced by variations of the airflow coming from the glottis, thus $p_i^a = \bar{u}/A$, where $\bar{u}$ is the varying component of the glottal air volume velocity and $A$ is the entry area of the vocal tract. Variations in the volume velocity are produced mainly by variations in the glottal area. Therefore, the approximation $\bar{u} = \bar{P} \cdot \bar{u}$ may be adopted, where $\bar{P} = \sqrt{2\Delta P/(k_1\rho_0)}$ is the mean particle velocity [7], $\Delta P = P_{st} - P_i^a$ is the static transglottal pressure difference and $\rho_0$ is the unperturbed air density. The vocal fold displacement may be written as $x(t) = \bar{P} + \xi(t)$, where $\bar{P}$ is an equilibrium position and $\xi(t)$ is the lateral displacement from that position. Then, the glottal area variation is $\bar{a}(t) = 2L\xi(t)$.

The reflected wave may be simply expressed as $p_i^r = -r\gamma p_i^a(t - v)$, where $0 \leq r \leq 1$ is a coefficient for wave reflection at the mouth, $0 \leq \gamma \leq 1$ is an attenuation coefficient that accounts for sound energy dissipation along the vocal tract, and $v = 2L/c$ is the time delay for the acoustic wave to travel forth and back the vocal tract tube, and $c$ is the sound speed.

Thus, the resulting pressure at the entry of the vocal tract is

$$P_i(t) = \bar{P} + \frac{2L\rho_0\gamma}{A}[\xi(t) - r\gamma\xi(t - v)].$$
4. Transglottal pressure at voicing onset

The above equations have an equilibrium position at \( \pi = \frac{P_t}{K} \). A standard stability analysis results in the oscillation threshold condition

\[
\Delta P = \frac{k_l(x_0 + \pi)}{2 \sin(\omega t)} \left[ \omega B = \frac{2\gamma L_{p0}(\pi)}{A} \sin(\omega t) \right] \tag{4}
\]

where \( \omega = 2\pi f_0 \).

To estimate the order of \( \pi \), let us consider an example with \( \bar{P}_t = 100 \text{ Pa} \) (from Fig. 1), \( K = 2 \times 10^6 \text{ N/m}^3 \) and \( x_0 = 1 \text{ mm} \). Thus, \( \pi = 0.05 \text{ mm} \ll x_0 \) and may be neglected. We also assume \( \tau \) small enough so that the approximation \( \sin(\omega t) \approx \omega t \) holds, and a value of \( k_l = 1.1 \) [7].

The reflection coefficient may be estimated using a low-frequency approximation for a flanged pipe in free space [11]. For \( f_0 \) values below 300 Hz and mouth area up to 5 cm², the approximation produces 0.993 ≤ \( r \) ≤ 1. The attenuation coefficient \( \gamma \) has been estimated in the range 0.2 – 0.8, for frequencies \( f_0 = 100 – 300 \text{ Hz} \) [12]. Letting \( \rho = 1.14 \text{ kg/m}^3 \), \( c = 356 \text{ m/s} \) (standard values for 100% moisture air at 37° C), and \( v = 1/(2F_1) \) (quarter-wave resonance tube), Eq. (4) becomes

\[
\Delta P = a - b \frac{\sqrt{\Delta P}}{f_0} \sin \left( \frac{\pi f_0}{F_1} \right) \tag{5}
\]

where \( a = 0.55 B x_0 / \pi \) and \( b = 80.7 \tau \gamma L x_0 / (A \pi) \).

In Eq. (5), the first term on the right side contains only parameters related to the vocal folds, and it represents the value of the transglottal pressure for voicing onset without the effect of the vocal tract. The second term adds the vocal tract effect. Note that it contains factor \( L x_0 / A \), which is the ratio of the glottal area (at rest) to the vocal tract entry area. The smaller the vocal tract area, the larger the influence of the vocal tract on the voicing condition.

For a numerical example, let us consider a male configuration with \( B = 1230 \text{ N/s/m}^3 \), \( x_0 = 1 \text{ mm} \), \( L = 1.4 \text{ cm} \), \( A = 2 \text{ cm}^2 \) and \( \tau = 1 \text{ ms} \) [7], a constant \( f_0 = 120 \text{ Hz} \), \( r = 1 \) and \( \gamma = 0.3 \). As Fig. 3 shows, the relation of the transglottal pressure for voicing onset vs. \( f_0/F_1 \) has a “U” shape with a minimum at \( F_1 = 2f_0 \). The figure also shows the range of values of \( f_0/F_1 \) for low and high vowels, extracted from the male data in Fig. 2. Since low vowels have higher \( F_1 \) values than high vowels, and both vowel regions are located at the left side of the minimum, the transglottal pressure is higher for low vowels.

In case of the measured data, \( f_0 \) is not constant and its variation must be included in the analysis. To this end, Eq. (5) was fitted to the data by minimizing the cost function

\[
F_{a,b} = \sum_{i=1}^{N} [\Delta P_i - \Delta P^*]^2 + 10N(DP - DP^*)^2 \tag{6}
\]

where \( N \) is the number of measures for each subject and DP is the difference of mean values of the transglottal pressure at voicing onset for low vs. high vowels (given in Table 1). In each term, the star represents measured values, which are subtracted from the respective theoretical values. The second term forces DP as a fitting target and penalizes the formation of a flat horizontal pattern (with \( DP \approx 0 \)). Factor 10\(N\) sets similar weights to both terms and was chosen by visual inspection of the results. The results are shown in Fig. 4 and Table 1.

Table 1: Difference between mean values of the transglottal pressure for low vs. high vowels

<table>
<thead>
<tr>
<th>Subject</th>
<th>Data (Pa)</th>
<th>Eq. (5) (Pa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>69.6</td>
<td>8.1</td>
</tr>
<tr>
<td>W2</td>
<td>132.2</td>
<td>0.0</td>
</tr>
<tr>
<td>W3</td>
<td>75.8</td>
<td>64.3</td>
</tr>
<tr>
<td>M1</td>
<td>85.9</td>
<td>83.3</td>
</tr>
<tr>
<td>M2</td>
<td>100.4</td>
<td>100.7</td>
</tr>
<tr>
<td>M3</td>
<td>96.2</td>
<td>105.9</td>
</tr>
<tr>
<td>M4</td>
<td>4.3</td>
<td>3.8</td>
</tr>
<tr>
<td>M5</td>
<td>41.4</td>
<td>42.1</td>
</tr>
</tbody>
</table>

In general, the theoretical model is able to produce a good approximation to both the data pattern and the observed pressure differences at low vs. high vowels (DP). In Fig. 4, the theoretical values deviate a bit from a “U”-shaped curve, due to variations in \( f_0 \) (note that \( \Delta P \) is a function of both \( f_0/F_1 \) and \( f_0 \)). Still, they follow the same pattern as in Fig. 3, with a minimum around \( f_0/F_1 = 0.5 \). The only exception is subject W2, where the model generates a horizontal flat pattern with \( b = 0 \) and consequently \( DP = 0 \). The data still suggests that a “U”-shaped curve could be accommodated; however, its minimum would be at some point with \( f_0/F_1 > 0.5 \), which Eq. (5) is not able to produce. A similar difficulty seems to exist for subject M4, for which the model also produces a flat pattern. In this case, the data seems to follow a “U” pattern with a minimum at \( f_0/F_1 < 0.5 \).
Nevertheless, the model produces a good approximation to the observed value of DP (see Table 1).

5. Conclusions

A simple theoretical model of the vocal fold oscillation suggests that the observed higher values of intraoral pressure at voicing onset for high vs. low vowels might be consequence of the acoustical coupling between the glottal source and the vocal tract. The performance of the model must be assessed considering the extreme simplifications introduced in the theory. For example, a fixed glottal configuration is assumed for each subject across all utterances, whereas the presence of the intrinsic $f_0$ effect already indicates that the assumption might not hold. Further, variations of the vocal tract shape between different vowels are characterized only through its first formant, regardless of its geometry. Those simplifications should be lifted in follow-up studies, to explore the extent to which the acoustical coupling alone explains the observed phenomenon.

Our results are in agreement with a recent theory developed by Titze on nonlinear source-tract coupling in phonation [13], as an extension to the classical linear source-filter theory. Nonlinear coupling effects, which also include frequency jumps and other voice instabilities, are caused by a dependance of the glottal flow on the vocal tract acoustics. Those effects typically occur when $f_0$ and its harmonics interact with the vocal tract formants. Our study shows that the nonlinear coupling may be a relevant factor in normal speech, even at low values of $f_0$.

6. Acknowledgments

This work was supported by CNPq/MCT (Brazil).

7. References