Waveform patterns in pitch glides near a vocal tract resonance

Tiina Murtola1, Jarmo Malinen2

1Department of Signal Processing and Acoustics, Aalto University, Finland
2Department of Mathematics and Systems Analysis, Aalto University, Finland
tiina.murtola@aalto.fi, jarmo.malinen@aalto.fi

Abstract
A time-domain model of vowel production is used to simulate fundamental frequency glides over the first vocal tract resonance. A vocal tract geometry extracted from MRI data of a female speaker pronouncing [i] is used. The model contains direct feedback from the acoustic loads to vocal fold tissues and the inertial effect of the full air column on the glottal flow. The simulations reveal that a perturbation pattern in the fundamental frequency, namely, a jump and locking to the vocal tract resonance, is accompanied by a specific pattern of glottal waveform changes.

Index Terms: speech production model, modal locking

1. Introduction
The source–filter theory of vowel production assumes that vocal fold vibrations occur independently of the vocal tract (VT) whose resonances modulate the vowel sound [1, 2]. Although this approach suffices for many purposes, some phenomena cannot be explained without filter-to-source feedback. Perturbations in vocal fold vibrations when the fundamental frequency $f_0$ is near the lowest formant $F_1$ are an example of such phenomena.

Although such perturbations have been reported in human speakers [3, 4, 5], excised larynges [6], physical laboratory models [7, 8, 9, 10], and computer simulations [4, 10, 11, 12], only a few speech modelling studies (namely, [4, 10, 12]) have given attention to the time-domain dynamics of $f_0$ glides near $F_1$, and only limited details are reported on waveform changes, e.g., amplitudes or pulse shapes, in the glides. Yet time-domain simulations are beingfit for studying the intrinsically time-domain speech production, and waveforms patterns can be used to compare speech models with in vivo measurements, e.g., [5], in greater detail than is possible with spectrograms.

The focus of this work is simulations of pitch glides over an isolated VT resonance. A VT geometry corresponding to [i] is used, and $f_0$ is varied over the interval $[150 \text{ Hz}, 310 \text{ Hz}]$ which contains the first VT resonance $f_{VT1}$. A low-order time-domain model comprising a biomechanical submodel of the glottis and acoustic submodels of the VT and subglottal tract (SGT) is used to investigate changes in glottal waveforms during these glides. The model contains two mechanisms by which the VT and SGT loads affect the glottis model. First, the effect of the air column inertia on the glottal flow is included. This approach is related, but not identical, to the level 2 interaction explained in [12]. Second, feedback from the VT/SGT acoustics directly to the vocal fold tissues is included. To our knowledge, glide simulation studies have not previously included this mode of feedback even though it is an important factor during the closed phase of the glottal cycle.

Here, conceptual difference between the lowest formant and the first VT resonance is neglected, i.e., $F_1 \approx f_{VT1}$.

2. The model
The model used in this work is described in detail in [13, 14, 15] and outlined briefly below.

2.1. Vocal folds
The vocal folds are modelled as wedge-shaped vibrating elements of thickness $L$ and length $b$ (Figure 1). The oscillations of the elements are described using equivalent mass-spring-damper systems (Figure 2) with equations of motion

$$\left\{ \begin{array}{l}
M_1 \ddot{W}_1(t) + B_1 \dot{W}_1(t) + K_1 W_1(t) = F_1(t), \\
M_2 \ddot{W}_2(t) + B_2 \dot{W}_2(t) + K_2 W_2(t) = F_2(t),
\end{array} \right. \quad (1)$$

where $W_j = [w_{j1} \ w_{j2}]^T$ are the displacements of $m_{j1}$ and $m_{j2}$ in the $y$-direction for the $j^{th}$ vocal fold ($j = 1, 2$) as shown in Figure 2. The load force pair $F_j(t) = [F_{j1}(t) \ F_{j2}(t)]^T$ is described below. The mass, damping, and stiffness matrices $M_j, B_j, K_j$, respectively, in (1) are

$$M_j = \begin{bmatrix} m_{j1} + \frac{m_{j2}}{4} & \frac{m_{j1}}{2} \\
\frac{m_{j1}}{2} & m_{j2} + \frac{m_{j1}}{4} \end{bmatrix},$$

$$B_j = \begin{bmatrix} b_{j1} & 0 \\
0 & b_{j2} \end{bmatrix},$$

$$K_j = \begin{bmatrix} l_1^2 k_{j1} + \frac{l_2^2 k_{j2}}{4} & l_1 l_2 (k_{j1} + k_{j2}) \\
l_1 l_2 (k_{j1} + k_{j2}) & l_2^2 k_{j1} + \frac{l_1^2 k_{j2}}{4} \end{bmatrix},$$

and the parameters can be found in Figures 1–2 and it has been assumed that $l_1 + l_2 = 1$. Symmetry is imposed on the vocal folds for this work by using $M = M_1$, $K = K_j$, and $B = B_1$, $j = 1, 2$, and by setting $F(t) = F_2(t) = -F_1(t)$. It is further assumed that $b = b_{j1} = b_{j2}$.

The glottal openings at the two ends of the vocal folds, denoted by $\Delta W_i$, $i = 1, 2$, are related to (1) through

$$\begin{bmatrix} \Delta W_1 \\ \Delta W_2 \end{bmatrix} = W_2 - W_1 + \begin{bmatrix} g \\ H_0 \end{bmatrix}. \quad (3)$$

The rest gap parameters $g$ and $H_0$ are shown in Figure 1.

When the vocal folds are closed, the load forces in (1) are produced by a nonlinear spring force with parameter $k_{H}$, accounting for the elastic collision of the vocal folds, and by the acoustic counter pressure $p_c = p_c(t)$ (see equation (11)) from the VT and SGT. Thus, the force pair for equations (1) during glottal closed phase is

$$F = F_H = \begin{bmatrix} k_{H} |\Delta W_1|^{3/2} - C_p p_c \\
C_p p_c \end{bmatrix} \quad \text{for } \Delta W_1 < 0. \quad (4)$$

The geometric coefficient $C_p = C_p(t)$ accounts for the moment arms and areas on which $p_c$ acts (see equation (12)).
2.2. Glottal flow and aerodynamic force

The main component of the air flow in the speech production system, to which the acoustic component acts as a perturbation, is assumed to be incompressible and one-dimensional, and to satisfy mass conservation and Newton’s second law. The glottal velocity \( v_0(t) \) of this flow component, through the control area \( hH_1 \) superior to the glottis, is then described by

\[
\dot{v}_o(t) = \frac{1}{C_{\text{iner}} hH_1} (p_s - R_g(t)v_o(t)),
\]

where \( p_s \) is an ideal driving pressure source, \( C_{\text{iner}} \) represents the inertance of the air column in the airways, and \( R_g(t) \) regulates pressure losses in the glottis and hence the part of the driving pressure that is not available to accelerate the air column [13, pp. 8–10]. This simplified approach takes into consideration the translational inertial effect of the accelerating air column not described by the linear acoustic equations. The inertial effect represents, by Newton’s third law, implicit feedback from the VT and SGT to the glottal velocity. Note that the glottal flow, i.e., volume velocity, is \( U(t) = hH_1 v_o(t) \).

The pressure loss in the glottis consists of two components

\[
R_g(t) = \frac{12 \mu H_1 L_9}{\Delta W_1(t)^3} + \frac{k_s H_1^2 v_o(t)}{2 \Delta W_1(t)^2} \]

The first term represents viscous pressure losses, and the second term accounts for entrance/exit effects (i.e., pressure loss at entrance and partial recovery at exit) which are not attributable to viscosity [15, p. 9]. The parameters \( L_9, k_s, \mu, \) and \( \rho \) are described in Table 2.

The air flow (5) produces an aerodynamic force on the vocal folds [15, pp. 9–10]. When the vocal folds are open, the load force in (1) is hence \( F = [F_{A,1} F_{A,2}] \), and, for \( \Delta W_1 > 0 \),

\[
F_{A,1} = \frac{\rho LH_1^2 v_0^2}{2 \cos^2 \phi} \left( -\frac{1}{\Delta W_1(\Delta W_2 - \Delta W_1)} \right)
+ \frac{1}{\Delta W_1} \ln \left( \frac{\Delta W_2}{\Delta W_1} \right) - C_p p_c, \quad \text{and}
F_{A,2} = \frac{\rho LH_1^2 v_0^2}{2 \cos^2 \phi} \left( \frac{\sin^2 \phi \Delta W_2 + \cos^2 \phi \Delta W_1}{\Delta W_1 \Delta W_2 (\Delta W_2 - \Delta W_1)} \right)
- \frac{1}{\Delta W_1} \ln \left( \frac{\Delta W_2}{\Delta W_1} \right) + C_p p_c,
\]

where \( \phi \) is the angle of the inclined vocal fold surface from the horizontal (Figure 1), and \( p_c \) is given by (11).

2.3. Vocal and subglottal tract acoustics

The total air flow in the airways has a component due to compressibility that can be treated using acoustics. One way of deriving the acoustic equations is to start with the equations of fluid and thermodynamics, and then remove all terms not required for describing a radiative transfer mode of energy and momentum, propagating at the speed of sound [16, Section 2]. The generalised Webster’s equation (see [17]) is used as acoustic loads to represent both the VT and the SGT

\[
\frac{1}{c^2(S^{(j)}(s))^2} \frac{\partial^2 \psi^{(j)}}{\partial t^2} + 2 \pi w(W^{(j)}(s)) \frac{\partial \psi^{(j)}}{\partial s} + \frac{1}{A^{(j)}(s)} \frac{\partial}{\partial s} \left( A^{(j)}(s) \frac{\partial \psi^{(j)}}{\partial s} \right) = 0,
\]

where the parameters \( \alpha, c, \) and \( j = ’V’ \) or ’S’ refers to the VT or the SGT, respectively. The solution \( \psi^{(j)} = \psi^{(j)}(s, t) \) is the velocity potential of the acoustic field. The VT/SGT geometry is represented by the area function \( A^{(j)}(s) \), the stretching factor \( W^{(j)}(s) \), and the sound speech correction factor \( \Sigma^{(j)}(s, t) \), all at distance \( s \in [0, L^{(j)}] \) from the glottis, where \( L^{(j)} \) is the length of the VT/SGT.

The following boundary conditions are used for (8)

\[
\left\{ \begin{array}{l}
ad^{(j)}(0, t) = -C_{\text{iner}}^{(j)} v_o(t), \\
ad^{(j)}(L^{(j)}, t) + \psi^{(j)} e^{\phi^{(j)}(s)}(L^{(j)}, t) = 0.
\end{array} \right.
\]

The first boundary condition couples the resonator to the glottal velocity (5) with scaling \( C_{\text{iner}}^{(V)} = h H_1 / A^{(V)}(0) \) and \( C_{\text{iner}}^{(S)} = -h H_1 / A^{(S)}(0) \). The normalised acoustic resistances \( \phi^{(j)} \) at the other ends are given in Table 1.

2.3.1. VT and SGT geometries

The VT geometry of a healthy 26 years old female corresponding to a prolonged [i] at \( f_o = 210 \) Hz is used. MRI data was acquired using the method described in [18], and the computational geometries were extracted using the custom software described in [19, 20]. The VT geometry and area function are shown in Figure 3, and the related VT geometry dependent parameter values are given in Table 1. The inertia parameter in the glottal velocity (5) is taken to be \( C_{\text{iner}} = C_{\text{iner}}^{(V)} + C_{\text{iner}}^{(S)} \).
Table 1: Physical and physiological parameters dependent on the VT and SGT geometries.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>VT [i]</th>
<th>SGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acoustic resistance ( \theta^{(i)} )</td>
<td>0.014</td>
<td>1</td>
</tr>
<tr>
<td>Termination area ( A^{(i)}(L^{(i)}) )</td>
<td>66 mm²</td>
<td>1000 mm²</td>
</tr>
<tr>
<td>Inertia parameter ( C_{\text{iner}}^{(i)} )</td>
<td>2820 kg/m⁴</td>
<td>1040 kg/m⁴</td>
</tr>
<tr>
<td>Tract length ( L^{(i)} )</td>
<td>136 mm</td>
<td>350 mm</td>
</tr>
<tr>
<td>1st resonance ( f_{R1} )</td>
<td>199 Hz</td>
<td>500 Hz</td>
</tr>
<tr>
<td>2nd resonance ( f_{R2} )</td>
<td>2798 Hz</td>
<td>1000 Hz</td>
</tr>
</tbody>
</table>

Figure 3: VT intersections extracted from MRI data and the resulting area function at distance \( s \) from the glottis.

A straight exponential horn is used as the subglottal area function

\[
A^{(S)}(s) = A^{(S)}(0)e^\varepsilon s, \quad \varepsilon = \frac{1}{L^{(S)}} \ln \left( \frac{A^{(S)}(L^{(S)})}{A^{(S)}(0)} \right),
\]

following [21]. The value for \( A^{(S)}(0) = 200 \text{ mm}^2 \).

2.4. The acoustic counter pressure

The feedback from the VT/SGT acoustics to vocal fold surfaces is realized as a product of the acoustic counter pressure \( p_c = p_c(t) \) and the geometric coefficient \( C_p = C_p(t) \) in vocal fold load forces (4) and (7) above.

The counter pressure is the resultant of sub- and supraglottal pressure components

\[
p_c(t) = Q_{pc}P \left( \frac{\partial \psi^{(V)}}{\partial t} - (0, t) - \frac{\partial \psi^{(S)}}{\partial t} (0, t) \right),
\]

where the parameter \( Q_{pc} \in [0, 1] \) enables scaling the magnitude of the feedback. This scaling is needed because determining the proportionality of the energy (i) in the underlying flow and (ii) in the superimposed acoustics is difficult to evaluate.

The geometric coefficient \( C_p \) is best understood in reference to the moment produced by \( p_c \) on the vocal fold wedge. Assuming \( p_c \) acts longitudinally on the vertical surface in Figure 1 and considering moments around the inferior edge of the wedge

\[
C_p = \frac{1}{2\pi} (H_1 - \Delta W_1)(2H_0 - H_1 - \Delta W_1).
\]

3. Numerical simulations

As a preparation of the pitch glides, the vocal fold mass matrix \( M \) is determined as detailed in [13, p. 14]. Long steady simulations are used to obtain an initial stiffness matrix \( K^0 \) that produces \( f_o \approx 150 \text{ Hz} \) and vocal fold damping that keeps the glottal flow amplitude unchanging, resulting in \( b = 0.01 \). Values for other constants are listed in Table 2.

In all simulations, the equations of motion (1) are solved using a fourth order Runge-Kutta method, the glottal velocity (5) using implicit Euler method, and the Webster’s equations (8) are discretised spatially using finite element method and temporally using Crank-Nicolson method (for further details, see [13, pp. 14–12 and 24–27]).

The pitch glides are achieved by changing \( K \) and \( p_s \).

\[
K(t) = 2.2^{2s/T} K^0 \quad \text{and} \quad p_s(t) = 2^{s/T} p_s^0.
\]

Changing \( p_s \) has only a small impact on \( f_o \) in these glides but it ensures that, in the absence of VT/SGT feedback, increasing \( K \) does not lead to significant changes in glottal pulse shapes (i.e., type of phonation).

Figure 4: A spectrogram of the glottal flow \( U \).

4. Simulation results

A spectrogram for the simulated glottal flow \( U \) in a glide with feedback level \( Q_{pc} = 0.1 \) is shown in Figure 4. A perturbation is clearly visible when \( f_o \) crosses \( f_{R1} \). This pattern is absent if the same simulation is carried out without VT feedback (i.e., \( Q_{pc} = 0 \)). Note that no subharmonics are observed in this glide.

Figure 5 (top panel) shows parameters which have been extracted from \( U \) pulse by pulse. Since subharmonics were not present, there was no ambiguity in the definition of a pulse.
The \( f_o \) trajectory shows the perturbation pattern in greater detail than Figure 4: a rapid increase (latter referred to as a jump for clarity) from approximately 180 Hz to \( f_{R1} \), a locking to a plateau at \( f_{R1} \) with small oscillations at the beginning, and a smooth release. Figure 5 (top panel) also shows the open quotient \( OQ \) and the closing quotient \( ClQ \), see [24, Fig. 4]. The lower panel of Figure 5 shows the amplitudes (i.e., maxima as glottis closes fully through the glide) of \( U \) and \( \Delta W_1 \), and full pulse shapes at five instants are shown in Figure 6. The jump in \( f_o \) (instant \( B \)) is accompanied by a peak in the \( OQ \) and \( ClQ \) values, i.e., a momentary increase in breathiness, and a minimum in \( U \) and \( \Delta W_1 \) amplitudes. The increased breathiness shows as a rounder and less skewed \( U \) pulse, and Figure 6 also displays more gradual opening and closing of the vocal folds and build-up and cut-off of the flow when compared to pre-perturbation instant \( A \).

At the beginning of the locking plateau, \( OQ \) and \( ClQ \) drop while the amplitudes of \( U \) and \( \Delta W_1 \) increase rapidly. By the instant \( C \) the pulse shapes have become very similar to the starting shapes (\( A \)). \( OQ \) and \( ClQ \) then continue to decrease (increased pressedness) and the amplitudes to increase through the locking event at fairly steady rates. The vocal fold oscillations amplitude reaches its maximum slightly after \( f_o \) starts to increase again (\( D \)). This is also when maximum pressedness of the phonation occurs. After instant \( D \), \( OQ \) and \( ClQ \) start to increase gradually while amplitudes decrease, until well away from the perturbation (\( E \)), the pulse shapes are very similar to the start (\( A \)). The similarity of the pulse shapes at the beginning and end of the glide was achieved by using the \( p_c \) control described above.

Figure 5: Top: \( f_o \) (black) on left axis, and open quotient \( OQ \) (solid gray) and closing quotient \( ClQ \) (dashed gray) on right axis. Bottom: Maximum \( U \) in each pulse (black) on left axis and maximum glottal gap \( \Delta W_1 \) in each pulse (gray) on right axis. Vertical lines indicate time instants shown in Figure 6.

**5. Discussion**

The changes in glottal waveform around the \( f_o \) jump and locking can be interpreted by considering energy in the system. The total energy in the model is the sum of the acoustic, flow, and vocal fold oscillation energy. At the jump, i.e., when \( f_o \) first reaches \( f_{R1} \), the energy contained in the vocal fold movements and the glottal flow is reduced to feed the resonating VT. Injected energy is needed to unlock the model to a frequency range different from the VT resonance. The glide controls (13) provide this energy injection into the glottal flow, and they also increase the storage potential of the vocal fold mass-spring system to accommodate a part of this energy injection at a frequency higher than the plateau value of \( f_o \). Finally, the eigenfrequency of the mass-spring system wins and \( f_o \) is released.

The pattern of waveform changes, accompanying the observed \( f_o \) locking pattern, is robust against variations in parameter values and some model modifications, such as changing the lip termination to a first-order impedance model approximation of the piston model [15, 25]. The existence of the \( f_o \) locking and the details of the \( f_o \) trajectory do, however, depend on the parameter values and model details. Besides the feedback parameter \( Q_{pv} \), the loss parameters (\( b, L_o, k_o \)) also play a critical role in determining the \( f_o \) trajectory. A low \( b \) value, for example, can reduce or remove the jump and locking, while a high \( b \) has the opposite effect. This supports the energy interpretation as \( b \) regulates how much of the injected energy is stored in the vocal fold movements and hence available to unlock the system.

At the jump, the increased \( OQ \) reduces the duration of strong acoustics–vocal fold feedback in the glottal cycle, helping the system to settle at the locking plateau. In contrast, higher \( OQ \) increases the duration of the glottal cycle where feedback from the VT/SGT acoustics directly to the glottal flow (level 1 interaction in [12], not included here) is at its strongest. Waveform variations have not been reported for glides using this interaction mode, so questions remain open regarding whether such models exhibits similar self-liming dynamic behaviour and which effect would be dominant in systems containing both feedback modes.

**6. Conclusions**

The simulations indicate that the feedback from resonator acoustics directly to vocal fold tissues can cause significant perturbations to vocal fold dynamics when \( f_o \) crosses a VT resonance. These perturbations are visible in both \( f_o \) and the glottal waveforms, indicating a special energy transfer pattern between the vocal fold vibrations, glottal flow, and VT/SGT acoustics.
7. References


