TrueTalk uses a sophisticated voice model to generate audio output from the pitch- and duration-annotated phoneme stream. We describe this model briefly here, including the various control parameters.

5.1 Vocal Tract Model

TrueTalk uses a *concatenative* method of synthesis. Pairs of fundamental sound units are strung together to form continuous utterances. The fundamental units are consonant-cluster/vowel and vowel/consonant-cluster combinations called *dyads*. These sound units are encoded. The complete inventory is called a *dyad table*. The dyad table contains on the order of 2000 dyads for English, each derived from the recording of a human voice. When converting a phoneme sequence to audio output, TrueTalk selects the appropriate dyads from the table, and glues them together before decoding.

TrueTalk uses *linear predictive coding* (LPC) to encode the dyads. One advantage of using an LPC representation is that it maps well onto a "lossless tube" model for the vocal tract. It is beyond the scope of this manual to discuss this model, but references are given in Appendix F. In the lossless tube model, features of the vocal tract — most importantly the size — appear as adjustable parameters.

Scaling the vocal-tract size in the voice model is important for making the synthesized speech sound as though it were generated by different speakers. Simply shifting the pitch and pitch range is insufficient; this merely sounds...
TrueTalk Reference Manual

as though the same speaker is talking in a different tone. Different people have different body sizes, and hence different vocal tract sizes.

**FIGURE 5-1 LPC Synthesis Model**

![LPC Synthesis Model Diagram]

TrueTalk’s LPC synthesis algorithm uses a glottal source to generate a train of impulse excitations, which are then filtered according to a set of vocal-tract parameters (i.e., the LPC parameters). In addition to being able to scale the vocal-tract size, TrueTalk also lets you control several glottal-source parameters, each of which contribute to what makes a particular voice sound unique.

Because the glottal-source part of the TrueTalk voice models the physiology of a human speaker, a number of features of the human voice appear as adjustable parameters. With TrueTalk you can adjust the open quotient, spectral tilt, aspiration, and glottal damping. All of these parameters apply to voiced sounds only — sounds for which the vocal folds are vibrating periodically. We describe each of these features briefly here.

*Open quotient* refers to the percent of the pitch period that the glottis is open. If we were to plot the flow of air through the glottis as a function of time, we would see something like Figure 5-2. Each hump of this plot corresponds to a single period of the vibration of the vocal folds, or a single “pitch period”. For part of this period the glottal airflow is zero, or the glottis is closed, and for part of this period the glottis is open. The open quotient is the percent of time for which the glottal airflow is non-zero.
Spectral tilt refers to a phenomenon that occurs in speech as we lower and raise our voice. Higher speaking volume corresponds to a higher airflow through the glottis, or higher humps in the above plot. The change in the airflow when the glottis closes is therefore more abrupt at higher volumes. This abruptness creates a “tenser” more “metallic” sound. At lower volumes the opposite is true. The less drastic fluctuations in airflow result in a “softer” sound.

Spectral tilt is often described in terms of frequencies. Abrupt changes in any signal means more high-frequency components in the Fourier representation. This is true for glottal airflow and speech sounds, too. In terms of the frequency, \( \omega \), of the speech waveform, spectral tilt is expressed as a function of some factor \( \alpha \) by which the power spectrum \( S(\omega) \) of the speech signal is transformed:

\[
S(\omega) \Rightarrow \omega^\alpha S(\omega)
\]

Aspiration, in the context of the TrueTalk voice model, refers to the amount of turbulent airflow through the glottis during voiced sounds. Airflow through the glottis is not entirely laminar. To make its voice sound more natural, TrueTalk models this effect by adding a turbulent term to the glottal source that is proportional to the instantaneous airflow. The turbulence is not strictly random but rather is correlated with the pitch period.

Glottal damping is a phenomenon related to air and energy “leakage” to the subglottal system. The lossless tube model of the voice normally assumes a rigid boundary condition at the glottis. The glottal damping feature of the
TrueTalk voice model approximates the effect of subglottal leakage which can greatly affect the acoustic analysis of the full vocal tract.

### 5.2 Voice Parameters

<table>
<thead>
<tr>
<th>Control</th>
<th>Synopsis</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>\uS&lt;gender&gt;</td>
<td>Set voice gender to male (m) or female (f)</td>
<td>m</td>
</tr>
<tr>
<td>\uS&lt;gender&gt;</td>
<td>Set the gender for the current minor phrase.</td>
<td></td>
</tr>
<tr>
<td>\uF&lt;scale&gt;</td>
<td>Scale the front of the vocal tract by 1/scale</td>
<td>65</td>
</tr>
<tr>
<td>\uf&lt;scale&gt;</td>
<td>Scale the vocal tract for the current minor.</td>
<td></td>
</tr>
<tr>
<td>\uB&lt;scale&gt;</td>
<td>Scale the back of the vocal tract by 1/scale</td>
<td>65</td>
</tr>
<tr>
<td>\ub&lt;scale&gt;</td>
<td>Scale the vocal tract for the current minor.</td>
<td></td>
</tr>
<tr>
<td>\gq&lt;fraction&gt;</td>
<td>Set the glottal-source open quotient.</td>
<td>0.55</td>
</tr>
<tr>
<td>\ga&lt;fraction&gt;</td>
<td>Set the glottal-source aspiration.</td>
<td>0.60</td>
</tr>
<tr>
<td>\gt&lt;factor&gt;</td>
<td>Set the glottal-source spectral tilt.</td>
<td>0.35</td>
</tr>
<tr>
<td>\gd&lt;factor&gt;</td>
<td>Set the glottal-source damping.</td>
<td>0.05</td>
</tr>
<tr>
<td>\gv&lt;value&gt;</td>
<td>Set the voiced excitation amplitude (volume)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

### 5.2.1 Gender

\uS<gender>, \uS<gender>. Set the speaking gender to either m (male) or f (female). The change takes effect for the entire current minor phrase. The first form, \uS, changes the voice gender for the current minor phrase only. The second form, \uS, changes the gender permanently (or until another gender-setting control sequence is issued).

Switching gender involves two different sets of changes. First, the voice-synthesizer module loads a different dyad table. Next, the voice-synthesizer module must load the pitch and glottal source parameters for that gender. These are used to further “massage” the LPC parameters before they are transformed into waveform data.
5.2.2 Vocal-Tract Size

\!uf<scale>, \!UF<scale>. Scale the front of the vocal tract by 1.0/<scale>. The default for <scale> is 65. Use f to apply a change to the rest of the minor phrase, use F to apply a change to the all subsequent text.

\!ub<scale>, \!UB<scale>. Scale the back of the vocal tract by 1.0/<scale>. The default for <scale> is 65. Use b to apply a change to the rest of the minor phrase, use B to apply a change to the all subsequent text.

Larger people generally have larger vocal tracts, men generally have larger vocal tracts than women, and adults have larger vocal tracts than children. A larger value for <scale> corresponds to a smaller vocal tract, and vice versa.

\uf50 \ub50 This produces a deep voice,
\uf100 \ub100 this voice sounds "squirrely,"
and this is the default voice.

5.2.3 Glottal-Source Parameters

The glottal-source control sequences each take a single, real argument in the range 0 to 1, with the exception of the voiced amplitude parameter, which accepts an argument within the range -1 to 1. Values which exceed these limits are set to the boundary values (e.g. \gt1.5 is changed to \gt1.0).

Changes take effect with the next word, and parameters remain as set until a subsequent control sequence is given for that same parameter.

\gq<frac>. Set the glottal open quotient, where the argument is the <frac> of the pitch period for which the glottis is open. The default open quotient is 0.55. A large open quotient yields softer phonation, with values near 1 producing a “breathy” sound. A smaller open quotient yields a tenser sound with less bass.

\ga<frac>. Set the aspiration, or the <frac> of turbulent airflow through the glottis during voiced sounds. The default aspiration parameter is 0.6. Larger aspiration values sound more “leaky” or “breathy”.
\textbf{TrueTalk Reference Manual}

\texttt{gt<factor>}. Set the spectral tilt. This scales the default power spectrum of the speaking voice by the frequency raised to the \texttt{factor} power. In other words, if we let $\alpha = \texttt{factor}$, $S(\omega) =$ power as a function of frequency $\omega$, the spectral tilt scales $S(\omega)$ according to the following rule:

$$S(\omega) \Rightarrow \omega^{\alpha} S(\omega)$$

If $\texttt{factor}$ is 0 then $S(\omega)$ is unchanged. If $\texttt{factor}$ is 1 then there is a high degree of spectral tilt. This is typical of a quieter, “whispering” voice. The default value for $\texttt{factor}$ is 0.35.

\texttt{gd<factor>}. Set the glottal damping to \texttt{factor}. Glottal damping refers to the boundary condition at the glottal end of the lossless-tube vocal tract. In the real vocal tract there is a finite energy loss at the glottis (i.e. it is not truly a fixed boundary). This condition affects the acoustic properties of the vocal tract. This also provides the model with a means of de-coupling the energy contributions to the speech sound of successive glottal pulses. The default for \texttt{factor} is 0.05. You should avoid extreme values (0 and 1).

\texttt{gv<value>}. Set the voiced amplitude to \texttt{value}. This parameter corresponds to the strength of the excitation used in producing voiced sounds. In effect, this parameter serves as a volume control and may be interpreted as such. The parameter may varied over the range -1 to +1 with a default value of 0. Generally, values above roughly 0.9 will result in some clipping - a harsh distortion possibly resulting in unintelligible speech. Values below -0.9 produce speech that could be characterized as a noisy whisper becoming increasingly unintelligible as the -1.0 limit is approached. Note that most of the perceived changes in speech amplitude are due to the changes in the voiced components. Therefore, this parameter may be interpreted as a volume control.