With TrueTalk you can control a variety of prosodic features. We describe these controls in this chapter. The discussion below focuses on the controls themselves, and not on the theory of prosody and intonation. We assume that you are already familiar with this topic. For more information on prosody and intonation theory you should consult Appendix F, “References”.

4.1 Introduction

One of the ways TrueTalk makes synthesized speech sound more natural is by mimicking human prosodic behavior. Features that you can control with TrueTalk include:

- lexical stress (or word accent).
- accents (or sentence stress),
- intonational phrasing,
- segmental duration,
- pitch and pitch range.

TrueTalk determines lexical stress, or accents, in the pronunciation algorithm. You can control pronunciation using the methods described in the previous chapter. There we also described how you can explicitly specify syllable durations when defining a pronunciation, or phonetic transcription (see page 3-2).
In this chapter we describe how to control accenting, intonational phrasing (pitch contours) and the pitch value and pitch range. We also describe the controls for speaking rate, and review how to control segmental duration. We begin by defining some basic, prosody-related terms.

4.2 Basic Concepts

TrueTalk processes text in major phrases. In general, a major phrase is a sentence as marked by a period, exclamation point or question mark. You can, however, force major-phrase boundaries with escape sequences (See “Process Flow Control” on page 6-40).

TrueTalk implements a partial version of Pierrehumbert’s model of English intonation [13,14]. The basic prosodic unit of Pierrehumbert’s theory is an intonational phrase. Intonational phrases in TrueTalk are called minor phrases, or simply minors. Intonational phrases are marked by tonal patterns that break an utterance into identifiable “chunks” of information. The tone, or pitch, of a phrase is measured by the fundamental frequency (F0) of the speaking voice.

While minor phrases are generally marked by punctuation — commas, semicolons and parentheses — they are not always. Consider the following example:

I just ate but I’m still hungry.

This sentence comprises two intonational or minor phrases. TrueTalk uses a stochastic algorithm based on the linguistic content of the input text to determine minor phrase boundaries.

We can force a minor phrase boundary in the above example by putting a comma after “ate”:

I just ate, but I’m still hungry.

We say that the above sentence now contains two orthographic minors. This is in contrast to the first form, which contains only a single orthographic minor. We make the distinction in TrueTalk because some intonation con-
trols apply to minors and some only to orthographic minors, before the phrasing algorithm has been applied. Commas, semicolons and parentheses are all orthographic minor delimiters.

TrueTalk applies a pitch contour to each minor phrase. As in Pierrehumbert’s theory, the most basic pitch features are H (high) and L (low) prominences. These are defined relative to one another: H is a higher pitch than L would be in the same context. With TrueTalk you can control both the absolute pitch and the pitch range.

The most perceptually prominent F0 peak in a minor is called the nuclear accent. TrueTalk lets you specify prominence magnitudes for the H and L points of the F0 contour relative to the nuclear accent.

An intonational phrase may consist of several intermediate phrases. An intermediate phrase contains a pitch accent — an intonational prominence assigned to a stressed syllable — and a phrase accent — the pitch trend from the last pitch accent to the end of the phrase. While the theory allows for a pitch accent for every stressed syllable, TrueTalk allows for only one intermediate phrase per minor phrase. You may, however, insert multiple floating phrase accents in a single minor phrase.

With TrueTalk you can specify explicitly five of Pierrehumbert’s six pitch accents: H*, L*, H+L*, L+H*, and L*+H. The “*” determines which pitch feature is aligned with the stressed syllable. You can specify the sixth pitch accent, H*+L, indirectly, as part of a full phrase contour.

Finally, every intonational phrase has associated with it a boundary tone (H or L). Boundary tones are indicated by a “%” suffix. The F0 contour for a minor phrase is therefore given by three tone markers. Examples include (L* H H%), the “interrogative” contour, or (H* L L%), the “declarative” contour.
4.3 Intonation: Accent and Phrasing controls

4.3.1 Accents

**TABLE 4-1 Accent Controls**

<table>
<thead>
<tr>
<th>Control</th>
<th>Synopsis</th>
</tr>
</thead>
<tbody>
<tr>
<td>!*&lt;tone&gt;&lt;prominence&gt;</td>
<td>Accent the following word; <code>&lt;tone&gt;</code> can be L*, H*, L+H*, L*+H, or H+L*; <code>&lt;prominence&gt;</code> can be any floating point number.</td>
</tr>
<tr>
<td>!!*&lt;tone&gt;&lt;prominence&gt;</td>
<td>Same as above but de-accent all subsequent words in the <strong>orthographic minor</strong>.</td>
</tr>
<tr>
<td>!-</td>
<td>De-accent the following word.</td>
</tr>
<tr>
<td>!c</td>
<td>Cliticize the following word.</td>
</tr>
</tbody>
</table>

\!*<tone><prominence>. Accent the following word with the optional tone and prominence. Tone may be L*, H*, L+H*, L*+H, or H+L*. If no tone is specified, TrueTalk uses L* for a yes-or-no question and H* otherwise. If no prominence is specified TrueTalk uses a default value of 1.

Prominence is a multiplier for the nuclear accent (the most prominent accent in the phrase), and can be any positive floating point number. Values greater than 2.0 are not recommended.

The following example puts a prominent pitch accent (twice the nuclear accent) on “me”:

```
You can give the money to \!*H*2 me!
```

This makes the statement sound like an answer to the question, “To whom can I give all this money?”

\!\!*<tone><prominence>. This behaves the same as \!* above, but also de-accents all subsequent words within the same **orthographic minor**.

Here we use the escape to emphasize incredulity:

```
I can’t \!! believe you would think that of me!
```
De-accent the following word.

Accenting and de-accenting controls are especially useful for controlling the pronunciation of complex nominals, noun-noun sequences such as “think tank”. The examples below illustrate how accenting controls can make complex nominals sound more natural.

<table>
<thead>
<tr>
<th>frying !- pan</th>
</tr>
</thead>
<tbody>
<tr>
<td>think !- tank</td>
</tr>
<tr>
<td>!* bite !- size !* pieces</td>
</tr>
<tr>
<td>!! school board meeting</td>
</tr>
</tbody>
</table>

Cliticize the following word. This de-accents the word and in general shortens its duration.

4.3.2 Phrasing Controls

<table>
<thead>
<tr>
<th>Control</th>
<th>Synopsis</th>
</tr>
</thead>
<tbody>
<tr>
<td>!_</td>
<td>Downstep every non-initial pitch accent in the orthographic minor.</td>
</tr>
<tr>
<td>!@</td>
<td>Downstep the next pitch accent.</td>
</tr>
<tr>
<td>!f&lt;tone&gt;&lt;prominence&gt;</td>
<td>Insert a floating phrase accent.</td>
</tr>
<tr>
<td>!p&lt;tone&gt;&lt;prominence&gt;</td>
<td>Set the phrase accent of the current minor.</td>
</tr>
<tr>
<td>!i&lt;tone&gt;&lt;prominence&gt;</td>
<td>Set the initial boundary tone of the current minor.</td>
</tr>
<tr>
<td>!b&lt;tone&gt;&lt;prominence&gt;</td>
<td>Set the final boundary tone of the current minor.</td>
</tr>
</tbody>
</table>

Downstep every non-initial pitch accent in the orthographic minor. This sequence lets you apply the (H*+L L L%) contour to a minor phrase (note that \!* does not recognize the H*+L argument). This compresses the pitch range for the high tone of every non-initial pitch accent. You can control the amount by which the pitch range is compressed with the \!{ escape (see “Pitch Variation” on page 4-9).

This control is appropriate for lists, where you want to have a “descending” contour. For example,
We have blackberries and blueberries and strawberries and loganberries.

Downstep the next pitch accent. This compresses the pitch range of the next accented word in the intonational phrase.

Insert a floating phrase accent. Tone and prominence values are optional. Tone may be H or L, and prominence may be any positive floating-point value, although values greater than 2.0 are not recommended.

Using this escape sequence you can approximate intermediate phrasing of the Pierrehumbert theory.

The following example illustrates how inserting a floating phrase accent can change the inflection of a question from being a choice between a variety of juices and no juice to being a choice between two juices:

Would you like apple juice \!fH2 or orange juice \!fL ?

Set the phrase accent of the current minor. Tone and prominence values are optional. The phrase accent controls the pitch from the pitch accent to the end of the phrase.

A high phrase accent conveys an “open-ended” or “expectant” tone. For example:

I did all my chores. \!! Now can I go out and play?

Set the initial boundary tone of the current minor. Tone and prominence values are optional.

Set the final boundary tone of the current minor. Tone and prominence values are optional.

Boundary tones complete the parameter set required to fully specify the F0 contours discussed by Pierrehumbert. The following examples use escapes to illustrate the default “declarative” (H* L L%) and “interrogative” (L* H H%) contours explicitly.
\!*L* May I \!*L* join you \!pH \!bH1.5 .
I can \!*H* speak for \!*H* myself \!pL \!bL .

Note that the interrogative contour can also be generated using the \!? escape described in the next section.

### 4.3.3 Specific Pitch Contours

#### TABLE 4-3 Pitch Contour Controls

<table>
<thead>
<tr>
<th>Control</th>
<th>Synopsis</th>
</tr>
</thead>
<tbody>
<tr>
<td>!?</td>
<td>Put a “yes-no question”/“interrogative” intonation pattern on the current minor.</td>
</tr>
<tr>
<td>!/</td>
<td>Put a “continuation rise” on the current minor.</td>
</tr>
</tbody>
</table>

In addition to specifying F0 contours explicitly by the pitch and phrase accents and the boundary tones, TrueTalk also lets you specify a contour for phrase with a single escape. The two contours are the interrogative contour (shown explicitly above in the \!b example), and the “continuation rise”.

\!?. Apply a “yes-no question” intonation pattern (L* H H%) — also referred to as an “interrogative” pattern — to the current minor phrase. This is equivalent to a judiciously placed \!*L* \!pH \!bH sequence.

TrueTalk normally uses an interrogative contour for sentences ending with a question mark, the exceptions being \!-questions. You can use this escape to force the interrogative contour on these sentences:

Where \!? are you going? Sit down and eat your dinner!

Note that the escape sequence need not be at the start of the phrase, it can be anywhere within the minor.

\!/ . Apply a “continuation rise” contour (L H%) to the current phrase.

This contour conveys that there is “more to come”. For example,

Next you should see \!/ a switch. Turn it on.

---

**Intonation: Accent and Phrasing controls** 4-7
You can also use this with the \!*L*+H pitch accent to generate a “rise-fall-rise” contour (L*+H L H%). This contour indicates uncertainty or incredulity. For example, the following response to the question, “Do you really like eggplant?”:

\!*L*+H Sort of \!/ .

indicates that the answer is qualified (e.g., “As long as it’s smothered in sauce.”).

4.4 Pitch Scaling

Besides being able to specify relative pitch values (H and L), TrueTalk also lets you control the absolute pitch and the degree to which it varies. These controls are described in this section.

Each of the parameters described in this section apply to the current phrase, be it a minor or a major phrase. The change in the pitch parameter is permanent (at least up to the next escape override) if you use the sequence \!{! instead of \!{.

4.4.1 Pitch Range

<table>
<thead>
<tr>
<th>Control</th>
<th>Synopsis</th>
<th>Defaults</th>
</tr>
</thead>
<tbody>
<tr>
<td>!(R&lt;value&gt;</td>
<td>Reference line of the current minor</td>
<td>Male: 96 Hz Female: 165 Hz</td>
</tr>
<tr>
<td>!(T&lt;value&gt;</td>
<td>Topline of the current minor</td>
<td>Male: 115 Hz Female: 195 Hz</td>
</tr>
<tr>
<td>!(B&lt;value&gt;</td>
<td>Baseline of the current minor</td>
<td>Male: 75 Hz Female: 140 Hz</td>
</tr>
</tbody>
</table>

The intonational tone markers H and L used extensively in the previous section are relative pitch values. They are defined with respect to a pitch reference line, which is an absolute pitch value given in Hz. High tones (H) have pitch values that are between this reference line and a soft upper limit called
the *topline*. The precise value is a function of the tone’s prominence. Low tones (L) have pitch values that are somewhere between the reference line and a soft lower limit called the *baseline*. Again, the exact value depends on the prominence of the tone. The difference between the baseline and the topline is often called the *pitch range*.

*Note that the pitch range parameters are different for the male and female voices. Changes you make to one voice do not affect the other. Also, permanent changes you make with the \!{! form of the control (see below) remain in effect for the given gender. This means that you can reset a pitch-range parameter, change the voice gender (see “Gender” on page 5-4), and then change the gender back, and the original change remains in effect for the original gender, regardless of how you may have changed the pitch range of the other gender.*

People often expand their pitch range at the start of a new paragraph when they are reading aloud. This signals a change of topic. TrueTalk mimics this behavior by expanding its pitch range by default at paragraph boundaries.

\!{R<value>, \!{!R<value>. Set the pitch reference line. The argument <value> is in Hz.

\!{T<value>, \!{!T<value>. Set the pitch topline. The argument <value> is in Hz.

\!{B<value>, \!{!B<value>. Set the pitch baseline. The argument <value> is in Hz.

In the following example we use a higher pitch and expanded pitch range to produce a more “animated” or “excited” sounding voice:

\!{R110 \!{T140 \!{B60 You wouldn’t believe all the homework I have!

4.4.2 Pitch Variation

Finally, TrueTalk lets you control the variation of the pitch within a sentence. Not all of these controls apply to all intonational contours.
Pitch Scaling

\(! {K < \text{ratio} >}\), \(! {F < \text{ratio} >}\), \(! {E < \text{ratio} >}\). Set the downstep ratio. The first form sets the parameter for the current major phrase, the second form sets it for the remainder of the input (or until reset again with this escape sequence). The argument <ratio> is a floating point value which determines the rate of downstepping, normally between 0 and 1 (0 and 100%).

To downstep means to compress the pitch range of high (H) tones for non-initial pitch accents. You can apply a downstep contour (H*+L L L%) to a minor phrase explicitly using the \(! _\) control sequence (see “Phrasing Controls” on page 4-5). The downstep ratio specifies the relative pitch range of all non-initial pitch accents, relative to the initial one. Thus a <ratio> of 0.6 compresses the pitch range to 60% of its default value.

The downstep contour often indicates that the speaker is opening a new topic. It is produced most effectively by increasing the pitch range from the default value and assigning a prominence of 1 to all the pitch accents whose range is to be reduced. For example,

\(! _\) \(! {T 150} \) \(! {K .5} \) Please \(! *H*1\) open your \(! *H*1\) textbooks to the \(! *H*1\) chapter on the Civil War.

\(! {E < \text{ratio} >}\), \(! {E < \text{ratio} >}\). Set the “final lowering” ratio. This compresses the pitch range for the last half second of a major phrase. Final lowering helps convey that the speaker is finished with a topic.

Final lowering applies to phrases that end with a low phrase accent, such as the declarative contour (H* L L%).

\(! {E < \text{ratio} >}\), \(! {E < \text{ratio} >}\). Set the “final raising” ratio. Final raising refers to an expansion of the pitch range at the end of a major phrase. This control is included for completeness.

### TABLE 4-5 Pitch Variation Controls

<table>
<thead>
<tr>
<th>Control</th>
<th>Synopsis</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>(! {K &lt; \text{ratio} &gt;})</td>
<td>Downstep ratio for the current major</td>
<td>60%</td>
</tr>
<tr>
<td>(! {F &lt; \text{ratio} &gt;})</td>
<td>Final lowering for the current major</td>
<td>90%</td>
</tr>
<tr>
<td>(! {E &lt; \text{ratio} &gt;})</td>
<td>Final raising for the current major</td>
<td>120%</td>
</tr>
</tbody>
</table>

4-10

Pitch Scaling
4.5 Durations and Rates

Just as pitch plays an important role in intonation, so does timing. TrueTalk lets you control both the average speaking rate and the duration of individual sounds. Chapter 3, “Controlling Word Pronunciation”, describes how to specify explicit durations for phonemes within explicit transcription entries. We expand on that slightly here.

4.5.1 Phoneme Duration

In Chapter 3 we describe how you can embed phonetic transcriptions in the input text stream. A transcription has the form:

```plaintext
\( \{ \text{orthography} \} \text{transcription} \) \)
```

Where transcription is a phonetic representation, constructed from the alphabet found in Appendix B, and orthography is arbitrary text, usually the orthographic representation for the word. As we noted in Chapter 3, “Controlling Word Pronunciation”, phonetic transcriptions can include explicit durations, given in centiseconds. For example,

```plaintext
\( \{ \text{Paris} \} \text{pe="rE[30] } \)
```

makes the last vowel (here an “ee” sound) 30 centisecond long, or about one third of a second.

You can apply a simple duration value, as illustrated in the previous example, to all phonemes except stops (p, t, k, b, d, g in the phonetic alphabet) and affricates (C and J — see Appendix B, “Phonetic Alphabet”). These require a bipartite description of the duration, where the first number represents the duration of the closure and the second number represents the duration of the burst. The single duration value for the transcription is constructed according to the following formula:

\[ \text{duration} = \text{closure} \times 1000 + \text{burst} \]

where all values are measured in centiseconds. Thus the following,

```plaintext
\( \{ \text{cheese} \} \text{ "C[7008]Ez} \)
```

Durations and Rates
renders “cheese” with the “ch” sound having a 7 centisecond closure and an 8 centisecond burst.

4.5.2 Silence Duration

<table>
<thead>
<tr>
<th>Control</th>
<th>Synopsis</th>
</tr>
</thead>
<tbody>
<tr>
<td>( &quot;*[duration]) )</td>
<td>Embed a (&lt;duration&gt;)-centisecond silence in the text.</td>
</tr>
<tr>
<td>!si&lt;duration&gt;</td>
<td>Add a (&lt;duration&gt;)-centisecond silence to the initial position of the current minor phrase.</td>
</tr>
<tr>
<td>!sf&lt;duration&gt;</td>
<td>Add a (&lt;duration&gt;)-centisecond silence to the final position of the current minor phrase.</td>
</tr>
</tbody>
</table>

TrueTalk provides two different mechanisms for inserting pauses in the synthesized speech. You can either embed the phonetic transcription for a silence (*), or you can insert silences at the beginning and end of minor phrases.

\( "*\[duration]\) \). The phonetic transcription for a silence with the specified \(<duration>\), where the \(<duration>\) units are centiseconds. Note that the primary stress symbols (*) is required.

\!si\<duration>, \!sf\<duration>. Add a \(<duration>\)-centisecond silence at the initial (\!si) or final (\!sf) position of the current minor.

In the following example we insert a 50-centisecond silence at the end of the first minor phrase (between “want” and “but”):

\!sf50 You can do it like that if you want but I prefer a different approach.

Notice that the control sequence can go anywhere within the minor phrase. Thus the above example is equivalent to,

You can do it like \!sf50 that if you want but I prefer a different approach.
The following also has the same effect:

\!si50 You can do it like that if you want but I prefer a different approach.

This example raises an interesting issue. Although the above sentence contains two minor phrases, it consists of only a single orthographic minor. TrueTalk uses a stochastic algorithm to determine phrase boundaries when internal punctuation is absent. This can sometimes make it difficult to determine a priori where minor-phrase boundaries will fall.

For example, you may want to render the above sentence as follows:

You can do it like that | if you want | but I prefer a different approach.

where here the “|” symbols mark the intonational phrase boundaries. There is currently no completely reliable mechanism for forcing phrase boundaries in input text for TrueTalk. For now you must rely on punctuation (e.g., commas):

You can do it like that, \!sf50 if you want, but I prefer a different approach.

In this example we put a 50-centisecond pause at the end of the parenthetical phrase, “if you want”.

4.5.3 Speaking Rate

Finally, TrueTalk lets you specify an average speaking rate.

<table>
<thead>
<tr>
<th>TABLE 4-7 Speaking Rate Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>!r&lt;rate&gt;</td>
</tr>
<tr>
<td>!R&lt;rate&gt;</td>
</tr>
</tbody>
</table>

You can vary the average speaking rate from roughly 85 to 355 words-per-minute using the control sequences from Table 4-7. The parameter <rate>
can either be a floating-point value greater than 0 and less than or equal to 2.0, or one of the following:

- slowest
- slow
- normal
- fast
- fastest

The default speaking rate (<rate> set to either normal or 1.0) is approximately 190 words per minute.

Numerical values for <rate> are admittedly not very intuitive. They are measured with respect to the default rate of 1.0, where a rate of 1.2 is 20% slower than the default and a rate of 0.8 is 20% faster.

Rate changes take effect from the place in the input text where the control sequence appears. For example,

\!r0.2 This is spoken quickly, \!r1.8 but this is spoken slowly. This is spoken at the "normal" rate.

You can therefore use the speaking rate controls to render individual words at different rates.

Sometimes saying \!r1.8 specific words \!r1 more slowly aids comprehension.