The RaP (Rhythm and Pitch) Labeling System

by Laura Dilley and Meredith Brown

1 Introduction and Overview

The RaP (Rhythm and Pitch) system is a method of labeling the rhythm and relative pitch of spoken English. The following is a tutorial for using this system. The tutorial assumes usage of Praat speech analysis software (Boersma and Weenink 2002), which is available for download at http://www.fon.hum.uva.nl/praat/.

The RaP system permits the capture of both intonational and rhythmic aspects of speech; this and other features distinguish it from other labeling systems. Four labeling tiers are used for annotating speech prosody. These tiers carry information about the syllabic organization and orthography of the speech (the "words" tier), its rhythmic structure (the "rhythm" tier), tonal patterns (the "tones" tier), and other information (the "misc" tier). This section presents an introduction to labeling using these four annotation tiers.

1.1 Steps to RaP labeling

Annotating the prosody of spoken utterances using RaP takes place according to the following three basic steps:

1. First, a Praat textgrid is created for a speech utterance from its soundfile containing the four labeling tiers listed above. This is accomplished by first creating a "words" tier by placing interval marks at syllable boundaries, as well as typing the orthography for each syllable in the appropriate interval. A blank "rhythm" tier is then generated from the sequence of syllables listed in the "words" tier. Finally, blank "tones" and "misc" tiers are created for the tonal and miscellaneous information.
2. Next, metrically prominent syllables and phrasal boundaries are labeled in the "rhythm" tier. This is done by listening to the relative prominence and phrasings of syllables in context while considering some simple heuristics for labeling speech rhythm to be described later in this guide.
3. Finally, tonal and other additional information is labeled in the "tones" and "misc" tiers. Tone labels include labels both for prominence-lending pitch movements ("pitch accents"), as well as tonal information at phrasal boundaries.

The following provides a brief illustration of how these labeling steps are carried out in practice. In the following examples, several rhythm and tone labels are introduced. Throughout this tutorial, the basenames of files illustrating examples will be given in double angled brackets, e.g. <<filename>>. To view the files, open the files "filename.wav" and "filename.Textgrid" in the Praat objects window, select both, and click 'Edit'. To bring up these example files automatically, open and run the Praat script *examples* and type in the basename of the file to be viewed. To bring up blank textgrids along with the soundfile for RaP labeling practice, run the script *practice* and type in the basename of the files to be viewed. To view both example and labeled practice textgrids simultaneously, run the script *view_examples*. Finally, to create addition blank textgrids, the script *make_practice_textgrid* can be used.

First, consider the example in <<anna1>>. Note that the orthography of the speech has been divided up into syllables on the uppermost labeling tier. The first step in deriving the RaP annotation is to label perceived prosodic boundaries and metrical prominences in the "rhythm" tier. Note that each short phrase sounds like its own prosodic "unit"; that is, the last syllable in each utterance marks a significant point of perceived phrasal disjuncture. As a result, each of these two syllables is labeled with "))" in the "rhythm" tier, which indicates that these syllables are at the right edge of a major prosodic phrase boundary. Moreover, the stressed syllables in each utterance are *metrical prominences*; that is, they are perceived as "strong beats" in context. (Metrical prominences will be discussed later, in the section on labeling rhythm and phrasal boundaries.) These syllables are labeled with "X" on the "rhythm" tier. In contrast, syllables which are metrically *non*prominent, such as the lexically unstressed syllables -*na*, are assigned no label. Finally, the "tones" tier carries labels which describe the tonal characteristics of metrically prominent and nonprominent syllables in the speech. In particular, syllables that are labeled as metrically prominent in the "rhythm" tier are assigned "starred" high and low tones, as indicated by the asterisk next to the tone label. Note that the ":"symbol in :H and :L indicates that these tones are the initial tones in each utterance. Moreover, the major prosodic boundary in each phrase is labeled with an "unstarred" tone lacking an asterisk: +L and +H, respectively.

Next, we turn to several additional examples, which illustrate other labels used in RaP. First, consider the examples in <<maria>>. The two speech examples in this file illustrate the two main metrical prominence labels which are distinguished in RaP. Metrically prominent syllables which are especially strong and salient perceptually in their contexts are assigned the label "X" for "major beat". On the other hand, metrically prominent syllables which are only of moderate strength perceptually are assigned the label "x" for "minor beat".

The examples in <<maria>> also illustrate several additional tonal labels used in the RaP system. In particular, these examples show the use of low or high "unstarred" tones (L+, H+, or +H) in metrically nonprominent positions, i.e., on syllables lacking a "x" or "X" label. We will ignore the ":" diacritic for now, returning to it later in Section 3.1. The choice of tone label depends on the relative pitch level with respect to an adjacent tone, as well as on the timing of the tone with respect to metrically prominent syllables. In the first example, unstarred L+ tones are indicated on metrically nonprominent syllables which have a locally low pitch relative to adjacent H* syllables. Similarly, in the latter two utterances, the unstarred high tones (H+ or +H) are indicated on metrically nonprominent syllables which have a locally high pitch relative to adjacent L* syllables.

In these examples the "+" diacritic on unstarred tones indicates *the relative position of a starred tone to the right or the left*. In the first example in <<maria>>, a "+" on the right side of each L tone indicates that the immediately following syllable has a starred tone (i.e., a H*). Similarly, in the second example a "+" on the right side of each H tone indicates that the immediately following syllable has a starred tone (i.e., a L*).[1] Finally, in the third example a "+" on the right side and left side of successive H tones indicates the relative position of the starred tone with respect to each unstarred tone. The

[1] If there are starred tones both to the right and to the left, the "+" is indicated on the right-hand side of an unstarred tone by default.

latter two examples thus illustrate a minimal pair contrasting the affiliation of an unstarred high tone with a rightward versus a leftward starred low tone. That is, the types and placements of tones are exactly the same in the second and third utterances, except that a high unstarred tone (H+) is situated on *Ma-* in the former case while the high unstarred tone (+H) is situated on *-ther* in the latter case. In general, unstarred tones which are medial in a phrase are constrained to occur in positions which are *next to* starred tones.

The examples in <<maria>> illustrate how tonal events within a phrase are described in the RaP system in terms of sequences of individual starred and unstarred tones. The two-tone sequences "L+ H*" and "L* +H" in RaP correspond to the bitonal pitch accents L+H* and L*+H of the ToBI transcription system (Silverman et al., 1992; Beckman and Ayers-Elam 1997). Moreover, the sequence "H+ L*" in RaP overlaps with the H+!H* pitch accent label in ToBI, which in turn is a notational variant of the H+L* accent proposed by Pierrehumbert (1980).

Labeling individual starred and unstarred tones in this way not only more accurately describes the timing and relative pitch levels of tonal events in English (cf. Ladd and Schepman 2003; Dilley, Ladd, and Schepman 2005; Dilley 2005), but it also permits symmetries in the English intonational system to be revealed in a manner that was not possible under ToBI. For example, consider that the first and second examples in <<maria>> are in fact "vertical mirror images" of one another. That is, if the first utterance is "flipped upside down" so that each low tone is replaced with a high tone and vice versa, then the result is the second utterance. ToBI labels do not permit this symmetry to be captured in the sequence of labels. Also, note that RaP makes it possible to describe the difference between the second and third examples in <<maria>> as simply a matter of a difference in the affiliation of an unstarred high tone with a rightward versus a leftward starred tone, respectively.

The examples in <<millionaire>> illustrate how RaP uses sequences of singleton starred and unstarred tones to capture distinctions which were captured in ToBI using the bitonal pitch accents L+H* and L*+H. The first example in <<millionaire>> illustrates a contour which is labeled in RaP as a sequence L+ H*; the L+ and H* tones in this case are associated with the nonprominent syllable *a* and the following prominent syllable *mil*, respectively. The second example in <<millionaire>> shows a contour which is labeled as a sequence L* +H; in this case, the L* and +H tones are associated with the prominent syllable *mil-* and the nonprominent syllable *lio-*, respectively.

The examples presented thus far serve to illustrate that there is a transparent relationship between tonal labels in RaP and the timing of pitches on target syllables in speech utterances. This consistent relationship is further shown in <<marilyn>>, which illustrates the phonetic characteristics associated with a leftward-aligning low unstarred tone. In this example, a +L tone aligns with respect to an immediately leftward starred H* tone; then there is a rise to a H* tone, followed by a fall to a +L tone. This leftward-aligning +L tone which *follows* the H* on *Mar-* can be compared with the rightward-aligning L+ tone which *precedes* the H* on *ri-* of *Maria* in the first example in <<maria>>. The two examples thus illustrate a near-minimal pair demonstrating the contrast between +L and L+. Note that in <<marilyn>> there is continuous interpolation in pitch and F0 between the +L tone and the following H*, indicating that the intervening syllables are not marked with phonological tones. The H* tones in this utterance have a

locally high pitch, and there is a steep fall at the end of the utterance to a low unstarred tone. Note that the symbol "[x]" is used on syllables which are heard as metrically prominent but which are lexically unstressed. This issue will be addressed later in the section on ambiguity and uncertainty in labeling speech rhythm.

The example in <<coffee1>> can be considered the "vertical mirror image" or inverse of the tonal pattern given in <<marilyn>>. That is, flipping <<coffee1>> upside down yields the intonation pattern in <<marilyn>>. In <<coffee1>>, a +H unstarred tone aligns with respect to a leftward low starred tone. The phonetic effect is to produce a pitch on the second (unstressed) syllable of *Allison* which is relatively higher than the leftward accented syllable.

A pair of unstarred tones can also occur in sequence on the "tones" tier. Consider, for example, the utterance in <<marion>>. Here, a leftward-aligning +L tone occurs just after a starred H* tone, while a following rightward-aligning H+ tone occurs just before a starred L* tone. In this context, each of the low tones reflects a locally low pitch associated acoustically with a local F0 minimum, while each of the high tones reflects a locally high pitch marked by an F0 maximum.

Two unstarred tones can also surround a single starred tone, as in <<american>>. In the first example in this file, the rightward-aligning L+ tone and the leftward-aligning +L tone both surround a single H* tone. These tones are associated with locally low pitches on the metrically weak syllables *Am-* and *ri-* of *American*. The second example in <<american>> gives the "inverse" of the first example, so that low tones are replaced with high tones and vice versa. Here, rightward- and leftward- aligning H+ and +H tones surround a single L* tone; these tones are associated with locally high pitches on the metrically weak syllables *Am-* and *ri-* of *American*.[2] These examples also illustrate another convention of the RaP system, namely that tones occurring on syllables marked as having questionable metrical prominence, x?, are unstarred rather than starred. The initial unstarred tones on the indefinite article *an* describe the initial high and low pitches occurring on the first syllables of each utterance. Note that there is no "+" diacritic indicated, giving just H and L, respectively; this is because the syllables associated with these tones are not adjacent to starred-tone syllables.

1.2 Labeling phrasal boundaries and equal tones

In the last section we introduced several aspects of rhythm and tone labeling. In particular, we focused on the basic distinction between metrically prominent and nonprominent syllables. Moreover, we presented only examples in which the tone labels alternated between high and low. Now that some aspects of labeling have been discussed, we turn to some additional points: conventions for labeling prosodic phrase boundaries and points of disjuncture, and the labeling of more complex tonal sequences. As discussed briefly above, prosodic phrase boundaries are labeled in the "rhythm" tier. The symbol ")" is used to indicate a small phrase boundary, while the symbol "))"is used to indicate a large phrase boundary. Uncertainty regarding the size of a boundary may also be indicated by the use of the "))?" diacritic. This marker indicates that the labeler is certain that a boundary is present, but is uncertain whether it is small or large. Another

[2] The tonal sequences L+ H* +L and H+ L* +H can be compared with the tritonal pitch accents L+H*+L and H+L*+H proposed for English by Grice (1995).

diacritic, ")?", may be used when a labeler is uncertain whether a boundary is present at all.

To illustrate a situation which warrants a phrase boundary label, consider the example in <<i_means1>>. In this example, there is a sense of a boundary (i.e., a small disjuncture) after the word *I*. This sense of disjuncture is captured in the "rhythm" tier through the labeling of a single parenthesis on the syllable corresponding to *I*. Note that the tonal pattern in <<i_means1>> is identical to that of the example in <<marilyn>> discussed earlier. The similarity in these two examples does not end with the tonal pattern; the rhythmic patterns in the two utterances seems to be quite similar as well. How can we explain the fact that the +L seems to give rise to a greater sense of disjuncture in <<i_means1>> than in <<marilyn>>, given that this sense of disjuncture is not related in any obvious way to intonational or rhythmic differences?

One difference between <<i_means1>> and <<marilyn>> is that they involve different morphosyntactic constituents. The +L occurs at a word boundary in <<i_means1>>, but it occurs in the middle of a word in <<marilyn>>. The RaP system captures the intuition that the prosodic structure of the two examples is similar by prescribing a similar transcription. RaP also permits the greater sense of disjuncture in the case of <<i_means1>> to be captured through the additional labeling of a small boundary in the "rhythm" tier. (In contrast, ToBI would likely ascribe the perceived difference in disjuncture to an *intonational* contrast between a L- phrase accent in the case of <<i_means1>> and an unstarred L+ tone (of a L+H* pitch accent) in the case of <<marilyn>>.)

A similar tonal pattern is given in <<i_means2>>. In this example, there is a fall in F0 after *I*, but the pitch then remains level rather than rising as in <<i_means1>>. The extended level-pitched region is captured by the use of an unstarred, rightward-aligning E+, which is referred to as an "equal tone". The sense of disjuncture resulting from the fall to the low +L at the word boundary in <<i_means2>> is captured through the labeling of a small phrase boundary, as indicated by use of ")" in the "rhythm" tier.

The example in <<i_means2>> can be compared with that in <<anna_lemay1>>. The utterance in <<anna_lemay1>> has a similar pattern of intonation and rhythm, supporting the similar RaP transcriptions that are afforded to each. The difference in perceived disjuncture again seems attributable to the different morphosyntactic properties of the utterances. This difference is captured through the use of ")" in <<i_means2>> but not in <<anna_lemay1>>.

The "inverse" tonal pattern to that in <<i_means2>> is shown in <<i_means3>>. Here, low tones are replaced with high tones and vice versa. The same tonal pattern as in <<i_means3>> is shown in <<anna_lemay2>>. Here again, the difference in perceived disjuncture seems to be attributable to the different morphosyntactic properties of the utterances. In both these pairs of utterances, ToBI would attribute the difference in perceived disjuncture between *I means insert* and *Anna Lemay* as arising from an intonational difference. In particular, ToBI would likely prescribe a phrase accent and intermediate intonational phrase boundary in order to describe the falling and rising intonation patterns at the right edge of *I* in the cases of <<i_means2>> and <<i_means3>>, respectively. In contrast, ToBI would probably prescribe an unstarred pitch accentual tone and no phrase accent in the case of <<anna_lemay1>> and

<<anna_lemay2>>. In this regard, a ToBI transcription appears to obscure a seemingly important similarity between these two kinds of examples.

Another example of a minor phrasal boundary at a word edge is given in <<armani12>>. Here, the final syllable of the name *Armani* coincides with a +H tone that aligns with respect to a leftward low starred tone. The small phrase boundary in <<armani12>> can be compared with the major phrase boundary occurring after *Armani* in <<armani10>>. In this example, the major phrasal boundary is marked by the presence both of a following pause, as well as a more complex tonal sequence at the end of the word *Armani*. Here, the complex phrase-related tonal movement is written as a sequence of two tones, +L H. Another example of this complex phrase-related tonal pattern is given in <<anna_incredulous>>.

We have already introduced the E+ label, which was illustrated in <<i_means2>> and <<anna_lemay1>>, among others. The E* label is illustrated in <<legumes1>> and <<legumes5>>. In <<legumes1>> an E* is situated on the metrically prominent syllable *vit-* of *vitamins*; this tone describes the level pitch which spans the region from *-gumes* of *legumes* through *vit-*. Similarly, in <<legumes5>> the E* also describes the level pitch which spans the region from *-gumes* of *legumes* up through *vit-*. Phonetically, the E* marks a metrically prominent syllable which has a pitch that is about the same as the immediately preceding syllable or syllables.

Next, <<legumes3>> illustrates how an utterance that begins with a level or monotone pitch contour is labeled in RaP. In this example, the syllables exhibiting level pitch are flanked by a sequence of equal tones, :E E*. Just as in <<legumes1>> and <<legumes5>>, the E* describes the level pitch which spans the region from the beginning of the utterance up through *vit-*. An :E tone is labeled at the left edge of this level region; RaP requires the initial syllable in each speech utterance to be marked with a tone. The :E is selected since the following speech material has a level pitch. (Recall also that an initial unstarred tone which is not adjacent to a starred tone is labeled with no diacritic, as discussed earlier for <<american>>.)

Another example of an utterance which begins with a monotone pitch is given in <<legumes4>>. As in <<legumes3>>, the utterance-initial level region is flanked by a sequence of equal tones: :E E*. The E* tone is aligned with the metrically prominent syllable *good*; consistent with this tonal choice, the pitches of all syllables up to and including *good* are at about the same level. The pitch of the immediately following syllable, *source*, is higher than that of *good*, a fact which is accounted for by the presence of a +H tone on *source*; this +H tone is attracted to the leftward E* tone. The syllable carrying this +H tone, in turn, participates in a subsequent region of level pitch spanning *source of vit-*. This level-pitched stretch is described by a E* at the region's right edge on *vit-*. Immediately thereafter, there is a rise in pitch across *vitamins*, ending in a high-pitched utterance-final unstarred high tone in the upper part of the speaker's range. There appears to be a monotonic F0 interpolation between the F0 on *vit-* and the end of *-mins*.

Yet another example of an utterance beginning with a monotone pitch contour is given in <<armani5>>. As in the *legumes* examples, the level region is captured by two flanking equal tones, :E E+. The E+ tone is associated with the final, non-prominent syllable in *Armani*. This syllable is followed by a metrically prominent syllable bearing a !L* tone which indicates a small step down associated with locally reduced pitch range. This syllable marks the beginning of another short region of level pitch on *knew the*; the

right edge of this region is again marked by a E+ tone. This repetition of tonal labels in sequence is noted in the "misc" tier through the use of parallelism markers: "(//" and "//)". The immediately following metrically prominent syllable marks the point of another small drop in pitch, which is marked with a !L* tone. Finally, there is a more significant drop in pitch across *millionaire* to the low unstarred tone at the end of the utterance.

Next, in <<legumes2>> there is again a long, low stretch of seemingly monotone F0 at the beginning of the utterance. However, closer listening to this portion of speech makes clear that the pitches of the syllables in sequence are not all the same. In particular, there is a small drop in pitch from the metrically prominent initial syllable to the second syllable, which is captured by the reduced pitch range labels !H* +!L. The pitch then stays level up through the indefinite article *a*, a fact which is captured with the E+ label. (Note that the reduced pitch range symbol is never used in conjunction with equal tone labels. Because equal tones always entail locally reduced pitch range, it would be redundant to use the reduced pitch range symbol on these tones.) The pitch of the metrically prominent syllable *good* is slightly higher; this is captured by a !H* tone label. The pitch drops slightly again on the next syllable, *source*, as captured by +!L. The pitch then stays level again through the preposition *of*, as indicated by a E+ marker. Note that the repetition in the sequence of labels is captured in the "misc" tier by the use of the parallelism labels, "(//" and "//)". Finally, there is a sharp rise in pitch on *vit-*, as captured by a H* tone, followed by a complex low-high tone sequence.

In all of the above examples which showed high or low tones, each high or low tone was followed by a tone of a different type. The final introductory example in this section introduces a contour in which a high tone or a low tone is followed by a tone of like type. That is, in <<mamalie_lemm>>, a high starred tone is followed by a high unstarred tone, and a low starred tone is followed by a low unstarred tone. Earlier examples illustrated that when high and low tones alternate, each high tone or low tone corresponds to an extremum (highest or lowest point) in pitch. However, when high and low tones do not alternate, this is not the case. In all examples, however, low and high tones indicate significantly lower and higher pitch, respectively, that the leftward tone.[3] This illustrates the fact that L* can show up either as a local F0 minimum in the context of a rightward high tone, or as a falling F0 contour in the context of a rightward high tone or low tone.

This concludes the present overview of labeling using the RaP system. In subsequent sections we will address labeling issues in more depth.

2 Labeling rhythm and boundaries using RaP

The perceived rhythm of speech is captured in RaP through the labeling of metrical prominences and boundaries in the "rhythm" tier. This section presents some detail regarding how this is accomplished. There are two goals for the upcoming discussion. The first is to aid new labelers in developing an awareness of speech rhythm by highlight examples of rhythmic speech. The second goal is to describe the labels for

[3] The exception is that in utterance-initial position low and high tones indicate lower and higher pitch with respect to the rightward tone, rather than the leftward tone.

indicating speech rhythm and phrasal boundaries in more detail and to discuss conventions for their use.

We will start by presenting some examples of rhythmic speech. Consider the example in <<pushups>>:

<<pushups>> You see Aaron doesn't like pushups.

In this example, the syllables *you, see, Aa-, does-, like* and *push-* are perceptually strong and thus can be labeled as metrical prominences using "X" or "x" in the "rhythm" tier. Note that it is fairly straightforward to tap to the rhythm created by these prominences. This may be in part because the prominences seem to occur at perceptually regular intervals in time. The phenomenon whereby prominent syllables seem to come at regular time intervals is called "perceptual isochrony." Perceptual isochrony is *optionally* labeled in the "misc" tier. It is indicated by placing the label "[pi" somewhere within the interval associated with the first beat and the label "pi]" within the interval associated with the last beat in the perceptually isochronous sequence. The square brackets thus enclose the region heard as regularly rhythmic.

Speech only intermittently sounds perceptually isochronous. Nevertheless, building an awareness of perceptual isochrony in speech is an important step in becoming a competent RaP labeler. It will therefore be useful to examine a few more examples of perceptual isochrony.

Consider next the example in <<understand>>. Here, the syllables *sim-*, *try-*, *get*, *un-*, and *stand* are all clearly strong. Moreover, they occur at approximately equal temporal intervals, even though careful listening indicates a slight speeding toward the end of the phrase. Note also that there is a correspondence between the rhythm and the pitch: all of the metrically prominent syllables except the last one have a low pitch, while the nonprominent syllables have a high pitch. Again, all metrically prominent syllables are labeled with "X" or "x".

<<understand>> I'm simply trying to get you to understand!

Another example of rhythmic speech comes from <<oj>>. The initial portion of <<oj>> is punctuated by staccato-like syllables which sound perceptually isochronous. The rhythm then seems to change, and a different perceptually isochronous rhythm emerges near the end of the speech utterance. To hear the regular rhythms of the syllables more clearly, try separately selecting and playing the region from *what do you mean...* up through *says*, followed by the region from *OJ* through the end of the utterance.

<<oj>> What do you mean when it says make into two tankers of OJ, what does that mean?

In upcoming sections, there will be additional opportunities to practice hearing and labeling speech rhythm.

2.1 Conventions for labeling speech rhythm

In general, labeling speech rhythm using the RaP system involves a combination of listening for metrically prominent syllables, together with applying a set of conventions for rhythm labeling. This section describes these conventions for rhythm labeling and provides some examples illustrating how they are applied. In general, these conventions capitalize on linguistic regularities in the rhythm of language to assist in determining which syllables to label as metrically prominent or not.

The guidelines for labeling metrical prominences in RaP are given below. These guidelines list several factors which affect which syllables are labeled as beats (or metrical prominences). These conventions include statements of preference for labeling an alternation of metrically prominent and nonprominent syllables, for labeling content words as metrically prominent and function words as metrically nonprominent, and for labeling syllables in a word or a phrase which are lexically stressed or phrasally stressed as metrically prominent. These conventions are followed by several examples illustrating their implementation.

I. ***Clash/lapse convention*****. Prefer a transcription in which metrically prominent syllables are separated by one or two non-prominent syllables.** It is widely recognized that in many languages prominent syllables tend to be separated by one or two nonprominent syllables (Halle and Vergnaud 1987, Hayes 1995). Thus, it makes sense for the RaP system to capitalize on this recognized alternation by encouraging labelers to encode rhythmic alternation in their labeling.

Note that the tendency to hear and label an alternation of metrically strong and weak syllables is mediated by speech rate to some extent. In locally slower speech, there is a greater likelihood that adjacent syllables may each be metrically prominent. In contrast, for locally fast speech, metrically prominent syllables might be separated by three or perhaps even four nonprominent syllables on rare occasions. Five or more nonprominent syllables in a row do not seem to occur, and this is not permitted in RaP labeling.

II. ***Lexical stress convention*****. For *polysyllabic* words, prefer a transcription in which (a) syllables with primary, secondary, or ternary stress are assigned metrical prominences ("x" or "X") while (b) unstressed syllables (including unstressed unreduced syllables) are not assigned metrical prominences.** This guideline refers to the fact that for words with more than one syllable, knowledge of the "dictionary stress" can help determine the locations of metrical prominences. Typically, syllables with both primary stress and/or secondary stress will be labeled as prominences. For example, both the first and third syllables in *Massachusetts* will typically be metrically prominent, since these correspond to the secondary and primary stressed syllables in the word, respectively. In contrast, the syllables *sa-* and *-setts* are unstressed and therefore should not be labeled as metrically prominent. Likewise, the *-o* in piano and *au-* in *autonomous* are unstressed unreduced syllables and therefore will typically not be metrically prominent.

For compound words such as *treehouse* and *campground*, both syllables will be listed in some dictionaries as stressed. However, in the RaP system, only the more prominent of the syllables of a compound word should be

labeled a metrical prominence. For example, the most prominent syllables in *treehouse* and *campground* are *tree-* and *camp-*, so these syllables will preferentially be labeled as beats, while *-house* and *-ground* will not be labeled as prominent.

III. ***Content/function word convention*. For *monosyllabic* words, prefer a transcription in which (a) content words are assigned metrical prominences and (b) function words are not assigned metrical prominences.** Content words include nouns (*John, room, answer, Selby*), "full" verbs (*search, grow, run, have*), adverbs (*really, completely, very, also, enough*), adjectives (*happy, new, large, grey*), numerals (*one, thousand, first*), interjections (*well, ugh, phew*), answers (*yes, no*), and so on. Function words include determiners (*a, an, the, that, more, much*), pronouns (*me, it, one*), possessives (*my, your*), prepositions (*in, on*), conjunctions (*and, but, or, when*), modal verbs (*should, can, must*), auxiliary verbs (*be, am*), particles (*no, not, nor*) etc. Content words will usually be beats, while function words will usually be nonbeats.

IV. ***Multiple-word phrase convention*. For multiple-word *phrases*, prefer a transcription in which the most prominent syllable(s) in the phrase are treated as beats and the least prominent elements of the phrase are treated as nonbeats.** For example, consider the phrases *stop sign* and *Main Street.* In both cases, the first monosyllabic word is typically stronger than the second monosyllabic word. Hence, the first word will be preferentially labeled as a beat, while the second word will be preferentially labeled as a nonbeat. In contrast, in the adjectival phrase *young man* the second monosyllabic word (*man*) is typically stronger than the first monosyllabic word (*young*). Hence, the second word will be preferentially labeled as a beat, while the first word will be preferentially labeled as a nonbeat.

It is important to emphasize that these conventions simply state guidelines regarding which syllables should be labeled as beats or not. The single most important determinant of the rhythm labeling in RaP is the labeler's perception of the global utterance rhythm, which can override any of these factors. However, it is important to strive for a transcription which best accords with the guidelines outlined above, especially since more than one rhythm can occasionally be "heard out" of speech material.

Now that we have presented the guidelines for labeling metrical prominences in speech, we turn to some examples of how these conventions can guide the labeling of speech rhythm. In particular, we will illustrate how the metrical prominences suggested by the guidelines are reconciled with the metrical prominences identified through listening to speech rhythm, in order to determine the appropriate rhythm labeling. First, consider the utterance in <<faster>>.

<<faster>> How about the faster one?

In this example, there is a clear sense of rhythm, even though this rhythm is not perceptually isochronous. The syllables that sound like beats are *how*, *-bout*, *fas-* and *one*.

Having determined which syllables are heard as beats, we can consider what the conventions have to say regarding which syllables should be labeled as beats. First, there are two polysyllabic words *about* and *faster*. The guidelines suggest that the primary stress syllables of these words, -*bout* and *fas*-, should be labeled as beats. The guidelines thus are in agreement with perception that these syllables are beats. Next, we can consider the guideline which applies to the monosyllabic words *how*, *one,* and *the*. All of these syllables are all function words, and there is a guideline indicating that these syllables should preferentially not be labeled as prominent. However, another guideline suggests that beats should be labeled on alternating syllables. If *how*, -*bout*, *fas*- and *one* are all labeled as metrical prominences, then prominences will fall on every other syllable. Perceptually, beats fall on each of these syllables. This percept, together with the provision that metrical prominences alternate, overrides the weak prohibition against labeling function words as metrical prominences. Thus, in this example, perception of where the beats agreed rather well with the placement of beats suggested by the guidelines.

For another example of how rhythmic perception and the labeling guidelines together influence rhythm labeling, consider <<justice>>.

<<justice>> Chief Justice of the Massachusetts Supreme Court.

Here, the sense of rhythm is quite clear perceptually: *Chief*, *Jus*-, *Mas*-, *chu*-, -*preme*, and *Court* are heard as beats. What do the guidelines have to say regarding where beats should be labeled, and how well does this accord with perception of the actual speech rhythm? The guidelines suggest that for polysyllabic words, the lexically stressed syllables should preferentially be labeled as beats. The lexically stressed syllables of polysyllabic words here are *Jus*-, *Mas*-, *chu*-, and -*preme*, all of which sound metrically prominent perceptually. This suggests that the guideline for labeling beats on polysyllabic words fits well with intuition about where the beats fall based on listening. Moreover, there are several monosyllabic words: *Chief*, *of*, *the*, and *Court*; one of the guidelines suggests that monosyllabic content words should preferentially be labeled as beats, while monosyllabic function words should preferentially be labeled as nonbeats. Here, the content words are *Chief* and *Court*, both of which are heard as beats, while the function words are *of* and *the*, both of which are heard as nonbeats. Comparing this outcome with the beats arrived at through listening, it can be seen that there is a good correspondence between the metrical pattern suggested by the guideline and the pattern suggested by listening. Finally, one additional guideline states a preference for an alternation of beat and nonbeat syllables, which essentially amounts to a mild dispreference for labeling adjacent beats on both *Chief* and *Jus*- as well as on both -*preme* and *Court*. However, in this case, the perception of rhythm and guidelines for labeling beats on polysyllabic and monosyllabic words together win out over the preference for alternation in beat labeling.

Next, consider the example <<power>>. Here, the speech is fairly slow and deliberate, and the speaker is hesitant possibly to the point of disfluency. Perhaps because the speech rate is rather slow, it seems possible to hear the monosyllabic function words *the*, *in*, and *that* as prominent, in spite of a general prohibition against monosyllabic function words being beats. The fact that the stressed syllables *gov*- and *pow*- from the polysyllabic words *government* and *power* sound like beats is consistent with the

convention that treats the lexically stressed syllables of polysyllabic words as probable beats. The slowness of the speech also makes it easier to hear adjacent syllables as being beats, including *the* and *gov-*, and *in* and *pow-*. Perception of rhythm and labeling conventions in this way are taken into consideration to give rise to a set of rhythm labels for this utterance.

<<power>> The government in power that...

Next, consider the examples in <<legumes3>> and <<legumes4>>, which exhibit a relatively faster rate of speech. In both of these examples, there is perceptual isochrony globally, which influences the perceived locations of the beats. Most all of these beats can be inferred from the guidelines, suggesting good agreement between rhythm perception and the linguistic knowledge-based conventions. In particular, the lexically stressed syllables *leg-* and *vit-* from *legumes* and *vitamins* are heard as beats, as is the monosyllabic content word *good*. The monosyllabic function words *are*, *a*, and *of* are not heard as beats, nor are the lexically unstressed syllables of the polysyllabic words *legumes* and *vitamins*. In each case, the presence of a metrical prominence or not on a syllable is consistent both with perception and with the labeling conventions. Only for one syllable do the labeling conventions conflict with perception: the monosyllabic content word *source* is not heard as metrically prominent. We suspect that the global perceptual isochrony plus relatively fast speech rate combine to make the listener "skip over" *source* and to tend not to hear it as a beat. Note, however, that *not* hearing *source* as a beat is consistent with the convention against labeling beats on adjacent syllables.

<<legumes3>> Are legumes a good source of *vit*amins?
<<legumes4>> Are legumes a *good* source of vitamins?

Next, the example in <<flipside>> illustrates further how perception of utterance rhythm and labeling conventions work together to help determine the rhythm labeling. First we will consider how the perception of rhythm in this example influences rhythm labeling. In this example, the rhythm can be heard at more than one "level" in this speech. On the one hand, it is possible to identify a very fast rhythm punctuated by *like-*, *-wise, you, have, flip-* and *side.* On the other hand, it is also possible to identify a slower rhythm which seems to include just *like-*, *you,* and *flip-*. Thus, perception suggests two possible placements for beats:

(a) Beats on *like-*, *-wise*, *you*, *have*, *flip-*, and *-side*
(b) Beats on *like-*, *you*, and *flip-*

Now we can consider how the guidelines can help select between these two possibilities. First, consider that the guidelines suggest that only the strongest syllables *like-* and *flip-* of the compound words *likewise* and *flipside* should preferentially be labeled as beats, while *-wise* and *-side* should not be labeled as beats. This gives rise to two "strikes" against (a), above. Next, we can see that the guidelines suggest that the function words *you, can*, and *the* should not be beats. *You* is heard as a beat under both perceptual interpretations, while *can* and *the* are not heard as a beat under either

interpretation. Thus, the prohibition against labeling function words favors neither (a) nor (b). Finally, consider that the guidelines suggest that the word *have*, which is a content word in this context, should be labeled as a beat. This gives rise to a single "strike" against (b). Taken together, these conventions suggest that (b) provides the more favorable placement of beats than (a), and beats are accordingly labeled in this example on *like*-, *you*, and *flip*-.

Another example of spontaneous speech with clear rhythm comes from <<i_believe>>. In this example, it is possible to fairly easily pick out the metrical prominences auditorily, in spite of the fast rate of speech. In particular, metrical prominences can be heard on the syllables -*lieve* and *box*- of the polysyllabic words *believe* and *boxcars*. Also, the monosyllabic words *I*, *long*, *aren't*, and *full* also sound like beats; some of these are function words, which one guideline suggests should not normally be beats. This example therefore illustrates how perception prevails over the knowledge-based guidelines in determining the rhythm of the utterance.

<<i_believe>> I believe so as long as the boxcars aren't full.

In sum, rhythm is labeled in the RaP system through a combination of listening to the metrical strength and timing of syllables, while considering the placement of beats suggested by knowledge-based labeling conventions. The examples in this section illustrate that in most cases, the locations of perceived beats in speech will be in good agreement with the locations of beats suggested by the guidelines. When there is a conflict, perception ultimately dictates the locations of metrical prominences in an utterance. In the upcoming sections we will focus on additional complexities in labeling speech rhythm.

2.2 Listening for speech rhythm

A few words are in order regarding the method for assessing speech rhythm perceptually. Rhythmic perception is widely recognized to be context-dependent. As a result, the rhythm of speech should always be judged with respect to a sizeable context. To judge speech rhythm, listen to lengthy syllable strings: four or five syllables long at the very minimum and preferably eight to ten syllables long. Determination of beats should never be made on the basis of listening to isolated syllables (unless the utterance consists of only a single syllable), and listening to two- or three- syllable long chunks to determine rhythm should also be avoided.

This context-dependency suggests a strategy for determining the rhythm of a long passage. To start out, listen to a long passage and try to determine which syllables are clearly beats; label these first. For hard passages, zoom in on successively smaller chunks of speech, listening in each case to as long a stretch as can be managed in order to determine which syllables are the beats. Continue to zoom in, listening to successively smaller portions of speech, until clear rhythmic percepts emerge. Determine the beats according to the conventions described above, together with perceptual judgments. Once a particular small passage of speech has been labeled for rhythm, continue labeling the speech rhythm by zooming out again to isolate a longer portion of the speech, *one which*

overlaps with the previous rhythm. This is important in order to preserve the context for perceptual judgments of the speech rhythm. By alternately zooming in and zooming out, it should be possible to obtain a complete rhythm transcription for a speech utterance.

One strategy which may occasionally be helpful is to tap to syllables which sound like beats. Tapping can be a useful a tool that may aid in beat identification. However, for fast speech or for speech in which the rhythm is unclear, it may be counterproductive to try to tap to the speech syllables. This is because tapping involves focusing on a motor activity – the coordination of finger movements with speech – rather than focusing strictly on the rhythmic percepts themselves. Thus, tapping should be used selectively only when it is deemed to aid in identifying beat syllables.

2.3 Ambiguity and uncertainty in labeling speech rhythm

The rhythm of speech is not always clear perceptually. This section addresses several types of ambiguity in speech rhythm. In particular, we will discuss how to recognize rhythmic ambiguity, as well as how to annotate it.

One type of ambiguity regards whether a particular syllable should be labeled as a beat. The ambiguity may arise because the perception of metrical prominence on a syllable is rather weak. For example, when there are two monosyllabic function words at the beginning of a phrase, the first of these words can sometimes sound like a weak or possible beat. The convention taken in such ambiguous situations is to label the first function word as "x?" for "possible beat". This occurs on the initial syllable sequences of <<slope>> and <<american>>.

Another type of ambiguity concerns uncertainty over *which* syllable or syllables are beats, as illustrated in <<kindergarten>>. Here, the rhythm is not at all clear perceptually during the first part of the utterance. *Class*- is clearly a beat, but it is not clear whether there are any beats in the four-syllable sequence *-rooms of the New*. It is likely that at least one syllable in this sequence is a beat, because of the convention which encourages labeling beats every two or three syllables (i.e., the clash/lapse convention). However, it is not until *Eng*- that a clear beat occur perceptually. How is the rhythm of the initial portion of this example to be labeled?

When a string of syllables occurs for which the rhythm is not clear perceptually, the rhythm labeling conventions should be relied on more heavily than usual to generate a sequence of rhythm labels. For example, the convention dealing with polysyllabic words indicates that the main lexical stress syllable *class*- in the word *classrooms* should preferentially be assigned a metrical beat while *-room* should be treated as a nonbeat. Another convention suggests that beats occur approximately every two or three syllables. If *class*- and *Eng*- are labeled as beats, then there should be a beat labeled on either *of* or *the*. If *of* is assigned a beat, then there is a single nonprominent syllable between *class*- and *of* and two nonprominent syllables between *of* and *Eng*-. Perceptually, it seems more natural to assign *of* a beat than *the*. Adopting this solution means that we must accept a monosyllabic function word as a beat, which in this case is permitted due to the fact that other factors converge to outweigh the prohibition against labeling such words as beats.

A third type of ambiguity relates to a situation in which the overall rhythmic context makes certain syllables sound metrically prominent which would not be expected to normally be so. For example, lexically unstressed syllables of polysyllabic words can

sometimes become beats. Those syllables which are classified as "unstressed unreduced" seem especially prone to being heard as beats in context. For example, consider <<heavy_rain>>. We already saw two such examples earlier, in <<marilyn>> and <<coffee1>>; whenever a lexically unstressed or unreduced syllable that is part of a polysyllabic word sounds like a beat, the solution is to use the label "[x]". The example in <<heavy_rain>> begins with a strongly perceptually isochronous rhythm. The metrical prominences fall in regular fashion on the lexically stressed syllables *heav-*, *poss-*, *-round*, and *sev-* of the polysyllabic words *heavy*, *possible*, *around*, and *seventy*, respectively, as well as the monosyllabic content words *rain* and *high*. However, there also seem to be beats on the lexically unstressed final syllables of *possible* and *seventy*. These syllables are lexically unstressed but belong to polysyllabic words; RaP prescribes the label "[x]" to indicate that these syllables unexpectedly sound like metrical prominences.

Another example of this phenomenon comes from <<dansville>>. In this example, the second syllable of *Dansville* sounds metrically prominent in context. However, this syllable is not a lexically stressed syllable. As a result, it should be labeled as "[x]" rather than "x" or "X".

At this point, we return to another type of ambiguity. This concerns two adjacent syllables, each of which sounds metrically prominent, where the speech is too fast for both of them to be labeled as beats. Such a case is illustrated in <<park1>>. To hear the ambiguity, isolate *go right out*. It is not clear whether *right* or *out* is more prominent. It seems possible to hear one or the other of the syllables as prominent but not both. Thus, in a frame of listening in which *right* is a beat, then *out* cannot be strong, and vice versa. How can we resolve this ambiguity and arrive at a consistent labeling convention?

The primary strategy in resolving an ambiguity of this type is to listen to more of the global rhythmic context. Based on this context, it may be possible to determine whether one or the other of the adjacent strong syllables sounds stronger and/or occurs at a moment in time which fits better with the overall rhythm. Suppose we isolate the section of speech from *It would be nice* through *back door.* Then upon multiple hearings it should become clear that *go* and *out* are "on the beat", while *right* falls "off the beat". The two adjacent syllables constituting the ambiguity are labeled in a manner which reflects both the ambiguity and also the resolution of the ambiguity. In particular, the less prominent syllable is assigned "x?", while the two syllables are labeled with angled brackets: "x?> <x". The angled brackets indicate the two syllables which participated in the rhythmic ambiguity.

Another example of two adjacent syllables sounding like strong beats comes from <<heat>>. In this example, the rhythm of the initial portion of the speech is not clear. Either *he* or *real-* in *really* seems prominent, but the speech is locally fast and so it does not seem appropriate to label both syllables as beats. How are we to determine the rhythm labeling for this example: should *he* or *real-* be labeled as a beat?

In a case such as this, the ambiguity arises due to the juxtaposition of a monosyllabic word produced with a full vowel (*he*) and the lexically stressed syllable of a polysyllabic word (*real-* of *really*). We can apply the Lexical Stress Convention and Content/Function Word Convention in order to determine the rhythm labeling in this case. The Lexical Stress Convention suggests that the lexically stressed syllables of polysyllabic words should typically be labeled as beats. Moreover, the Content/Function

Word Convention suggests that monosyllabic function words should typically be labeled as nonbeats. In other words, the lexically stressed syllable *real-* of the polysyllabic word *really* should be labeled as a beat; this syllable is quite strong, so we can label it as X. Moreover, the monosyllabic function word *he* should be labeled as a nonbeat, x?. Finally, angled brackets are used to indicate the adjacent metrical ambiguity, so that we obtain x?> <X as the final labeling for *he real-*.

Another example of an ambiguity related to which of two adjacent syllables is stronger comes from <<stretch>>. Here, there is ambiguity in whether *you* or *stretch* is more prominent. Unlike some other cases, listening to a larger portion of the global rhythmic context in this example does not seem to help.

A guideline which is relevant to determining the rhythm labeling here is the convention for labeling monosyllabic words. Here, *stretch* is a content word, while *you* is a function word. Thus, we should preferentially label *stretch* as a beat and *you* as a nonbeat. Consistent with this, the word *stretch* does in fact seem to be slightly stronger perceptually compared with *you*.

Another example of metrical ambiguity involving two adjacent strong syllables comes from <<two-inch>>. In this example, the syllables *two* and *inch-* are each strong, but it does not seem correct to label them both as beats. Several considerations suggest that we should select a labeling in which *inch-* is labeled as a beat and *two* is labeled as a nonbeat. First, listening to the global context suggests that *inch-* comes at a point in time that makes it seem more prominent than *two*. This choice of relative metrical prominence also means that every pair of beat syllables is separated by one or two nonbeat syllables. In contrast, if *two* were to be labeled as a beat and *inch-* as a nonbeat, then two consecutive syllables would be labeled as beats (*-bout* and *two*), and there would be a sequence of three syllables (*-ches it's like*) labeled as nonbeats. Thus, labeling *inch-* as the beat results in more even spacing of beats throughout the utterance. Finally, a convention was presented earlier that when a rhythmic ambiguity is created through the juxtaposition of a monosyllabic word and the lexically stressed syllable of a polysyllabic word, a labeling should be preferred in which the lexically stressed syllable is labeled as a metrical beat.

Another example of metrical ambiguity in adjacent syllables comes from <<bruins>>. In this example, the rhythm is clear at the very beginning of the utterance and at the end, but it is somewhat obscured in the region of the phrase *face off*. At this point in the utterance, there are conflicting contextual rhythmic cues to prominence. If *the Boston Bruins face* is played by itself, then *Bos-*, *Bru-* and *face* each sound like beats. However, if *off against the Buffalo* is played by itself, then *off* sounds like a beat. Here, we utilize the convention relating to multiple-word phrases. The phrase *face off* is used in this context as a verb, so that *off* is stronger than *face*. Thus, *off* should be labeled as a beat while *face* should be labeled as a nonbeat.

A particularly tricky situation is when the pattern of metrical prominences suggested by the global rhythmic context conflicts with lexical stress. For example, consider the example in <<asylum>>. In this example, the initial portion of the phrase sets up a rhythm with beats on *can't* and *sy-* of *asylum*. These two strong beats set up a rhythm, and the listener's expectation is that the rhythm will continue. In particular, the listener expects that the temporal interval separating the beats on *can't* and *sy-* will be matched by a downstream beat at the expected moment in time, thereby creating an equal

interval. As it turns out, the lexically unstressed syllable *be-* of *because* occurs at the expected moment in time; as a result, *be-* sounds unexpectedly prominent. However, it was stated earlier that lexically unstressed syllables of polysyllabic words cannot typically be beats. The solution is to treat this as a case of adjacent prominence ambiguity and to assign the beat to the lexically stressed syllable *-cause* in *because*.

Finally, we will discuss one last type of ambiguity, which concerns uncertainty related to the level of strength or prominence of a syllable. Some syllables will clearly be of exceptional strength, while other syllables will clearly be of only moderate strength. However, it will not infrequently be the case that the degree of prominence of a particular syllable seems somewhat unclear, leading to ambiguity in whether "X" or "x" should be labeled on a particular syllable. In this case, a good label choice is "X?" This label indicates that the labeler is certain that a syllable is a beat, but is not certain whether it is exceptionally strong or only moderately strong.

We will pause here to summarize the different types of rhythmic ambiguity discussed in this section and how they are labeled. First, there can be ambiguity related to whether a particular syllable should be labeled as a beat or not, such as a monosyllabic function word at the beginning of an utterance. Words of this sort can be labeled with "x?" to indicate possible beats. Second, there can be ambiguity relating to which syllable or syllables are beats when the rhythm is not perceptually clear. In this case, the rhythm labeling guidelines should be relied on more heavily to determine the appropriate labels. Third, syllables which would not normally be labeled as beats can sound metrically prominent in context but, such as the unstressed unreduced syllable *-lyn* in *Marilyn*. Such syllables can be labeled with "[x]" to indicate that they are perceived as metrically strong. Fourth, there can be ambiguity in which of two adjacent strong syllables should be labeled as a beat. Here, global rhythmic percepts, together with labeling conventions, are used in assessing which of the syllables was likely to be more prominent. The more prominent syllable of the two syllables is assigned "x", while the less prominent syllable is assigned "x?", and angled brackets are used to indicate the ambiguity. Finally, a syllable which is judged to be metrical prominence but whose level of strength is unclear are labeled with "X?".

Now that we have considered the ways that ambiguities are labeled in RaP, we turn to the issue of how phrasal boundaries are labeled.

2.4 Labeling phrasal boundaries in fluent and disfluent speech

A major characteristic of the RaP system is its emphasis on *perception* in labeling phrasing and disjuncture in speech. Words or syllables which are perceived as "final" in a phrase, or which are identified as positions of disjuncture are assigned specific labels in the "rhythm" tier. In particular, positions which are perceived as points of major or significant disjuncture are indicated by labeling a double parenthesis "))" on the target syllable; these positions are referred to as "major phrasal boundaries." In addition, positions which are perceived as points of minor disjuncture are indicated by labeling a single parenthesis ")"; these positions are referred to as "minor phrasal boundaries." Several examples illustrating the use of these boundary markers were provided earlier, and additional examples will be presented later in this section.

It is important to note that there is little dependency in the selection of phrasal labels on the "rhythm" tier and the selection of labels on other tiers. In particular, the selection of a phrasal marker on the "rhythm" tier does not oblige the user to label a tonal marker on the "tones" tier. In this regard, the RaP system differs from ToBI, which requires a tonal label to be selected every time that a phrasal boundary marker is labeled. This lack of dependency between the "rhythm" and "tones" tiers is consistent with the fact that there are a number of phonetic characteristics which can lead to a sense of pausing or disjuncture, including pausing, segmental lengthening, glottalization, etc. A pitch obtrusion may or may not coincide with the perceived end of a phrase; likewise, a tonal label may or may not be selected on the "tones" tier when a phrase boundary is labeled in the "rhythm" tier. Conventions for labeling tones are discussed later in this manual.

There are two labels for marking uncertainty with respect to phrasal boundaries in the RaP system. The label "))?" is used when the labeler is certain that there is a greater than normal degree of disjuncture between syllables but is uncertain regarding the size of the boundary. Moreover, the label ")?" is used when the labeler is not certain whether a phrasal boundary is present between two syllables.

We now turn to the issue of other kinds of events which can induce a sense of phrasal disjuncture, namely disfluencies, hesitations, cutoffs, and restarts. In the RaP system, disfluencies which induce a sense of disjuncture are labeled as major or minor phrases in the "rhythm" tier using the "))" or ")" symbols, respectively. In addition, such events are distinguished from fluent pausing and phrasal boundaries through the use of flags in the "misc" tier. The marker "dis" is used in the "misc" tier for a generalized disfluency, while "cut" and "hes" are used in the case of a syllable or word which is cut off or in the case of a hesitation, respectively. In addition, "res" is used to mark a restart. More tha none marker can be used in the "misc" tier to describe a disfluent speech event, e.g. "dis/hes". The following presents some examples of disfluent speech and how it is labeled in the RaP system.

The example in <<avon>> illustrates two properties which frequently appear in disfluent speech. First, disfluencies are often associated with an inappropriate sense of pausing. In this example, there is a disfluent pause after the word *from* which perceptually sets this word apart from the following speech material, thereby inducing the sense of a boundary after the word. To capture this sense of disjuncture, a major phrase boundary is labeled after the syllable corresponding to *from* in the "rhythm" tier. The disfluent syllable is flagged with the "dis" marker in the "misc" tier. Second, disfluencies can sometimes induce a word which would not normally be prominent to sound strong or accented. Here, the function word *from* sounds unexpectedly prominent in context. Because of its perceptual prominence in this context, *from* is labeled as a beat in the "rhythm" tier. RaP permits the capture of both the sense of disjuncture and the sense of unexpected prominence associated with the disfluency through coordinated labels across several labeling tiers.

Disfluent speech will often exhibit another somewhat surprising property. That is, the overall rhythmic structure can seem to be well-formed in the vicinity of a disfluency. The rhythm created by disfluent syllables may even be perceptually isochronous. The combination of a lack of fluency together with a well-formed rhythm is illustrated in the example <<graft>>. There are several disfluencies in this utterance, yet there is a clear

rhythm which pervades the utterance. In *oh the there's a joke*, the syllables *oh*, *there's*, and *joke* form a perceptually isochronous rhythm, and each syllable is labeled as a beat. The disfluency on *oh the* induces a sense of a small boundary, as does the slight hesitation after *there's a*. These disfluencies and hesitations are noted by a combination of "dis" and "hes" markers in the "misc" tier. In spite of these disfluencies, the locations of metrical beats are clear, and there seems to be a connected sequence of metrical beats which emerge from the syllable sequence in this example.

Another example of how disfluencies and hesitations can be associated with a clear sense of well-formed rhythm comes from <<meadow>>. The initial portion of this example is associated with hesitation, as well as use of the filler word *um*. Nevertheless, there is a clear sense of rhythm, and the syllables *go*, *um*, *rec-* and *-cross* even form a perceptually isochronous sequence. (Note that labeling "dis" on filler words such as *um* and *uh* is optional.)

A further example of how disfluency interacts with rhythmicity is atlanta>>. The initial portion of this example is highly disfluent, with several words being cut off. (Note that these are labeled in the "misc" tier using the marker "cut/dis".) Somewhat remarkably, the sequence of syllables associated with the rather striking disfluency in *yes I would l- uh like the information* corresponds to a perceptually isochronous sequence; that is, the syllables and fragments *yes*, *I*, *l-*, *like*, *in-*, and *mat-* are metrically strong and exhibit temporal regularity. This example also illustrates that a syllable fragment (the hesitated *l-* in the intended word *like*) can itself be a beat.

Labeling practice. Rhythm, disfluency, and rhythmic ambiguity.
Label the rhythm and phrase boundaries for the following examples, as well as any perceptual isochrony. Do not transcribe tones.
<<engine>>
<<escaped>>
<<insects3>>
<<dukakis>>
<<nashville>>
<<encroach>>
<<elephant>>
<<either_or>>
<<hyannis>>
<<park2>>
<<nonstop>>
<<business>>
<<musicians>>
<<capote>>

---

3 Labeling tones and other information

In the introduction, we showed that the RaP system uses several tonal symbols to capture the phonologically significant up-and-down patterning of the tones in speech. In

this section we consider in more depth a variety of issues related to labeling information in the "tones" tier. The first issue to be addressed concerns how the tonal markers introduced earlier relate to phonetic and perceptual characteristics associated with intonational variations in speech. The second issue concerns conventions on the usage of these tonal labels. Both of these issues will be addressed in this section, and additional examples will be presented illustrating the use of RaP tonal labels.

3.1 Perceptual and acoustic-phonetic characteristics of tones and tone sequences

What are the important phonetic and perceptual characteristics associated with tonal markers in the RaP system? The most important phonetic characteristics are the perceived patterns of relative pitches of syllables, in conjunction with the perceived metrical structure. Specific tonal markers in RaP are associated with particular phonetic and perceptual characteristics. Conveying the phonetic and perceptual characteristics associated with these tonal markers will lead to an elaboration of the labeling conventions that have been conveyed so far.

In RaP, there is a clear and consistent relation between a tone label and the pitch of the associated syllable relative to other syllables. Each RaP label is selected to reflect the relative pitch level of a tonally-marked syllable with respect to the immediately preceding tonal context, given some rhythmic structure.[4] In the following we consider in more depth how RaP labels are selected and which phonetic properties they are associated with.

First, we will consider how the choice of high, low, or equal tones relates to perceived relative pitch level, for tones in *non-utterance-initial position*. The starred tones H*, L* and E* describe metrically *prominent* syllables which have a pitch that is respectively higher than, lower than, or equal to that of the previous syllable associated with a tone (i.e., the previous tone-associated syllable). Similarly, H, L and E unstarred tones describe metrically *nonprominent* syllables which have a pitch that is respectively higher than, lower than, or equal to the pitch of the previous syllable associated with a tone. (We will ignore the "+" diacritic for the moment).

We note that all transitions or "interpolations" between adjacent tone pairs are monotonic. Thus a (non-utterance-initial) *high* tone always indicates that the associated syllable has a *higher* pitch than the preceding tone-associated syllable, as well as a higher pitch than every syllable between the two tones. Similarly, a *low* tone always indicates that the associated syllable has a *lower* pitch than the preceding tone-associated syllable and every syllable in between. Finally, an equal tone indicates that the associated syllable has a pitch which is *equal* to that of the preceding tone-associated syllable, as well as to every syllable between the two tones. It may be useful at this point to review the tonal transcriptions of some examples presented earlier with these phonetic characteristics in mind: <<maria>>, <<millionaire>>, <<marilyn>>, <<marion>>, <<i_means1>>, etc.

Note that just as for speech rhythm, it is important to evaluate the perceived relative pitches of syllables *with respect to their context.* In particular, at least one syllable of context to the left and right of the region of speech to be evaluated, and preferably more, should be listened to when deciding on a tonal transcription. This is

[4] Note that utterance-initial tones have a consistent relative pitch level with respect to the immediately *following* syllable(s).

because the perceived pitch of syllables can be influenced by which portion of the file is played. The process of isolating and playing some portion of the speech material introduces discontinuities in the signal which can make some regions of the speech sound more salient than they would in context. Because RaP aims to identify the phonologically salient pitch events that participants would hear during listening situations, it is very important when evaluating RaP labels within the global context of a speech stream.

Next, we can consider the phonetic characteristics associated with tones in *utterance-initial* position. For tones in this position, there is no preceding tonal material with which to compare the pitch level of an utterance-initial tone. As a result, tone labels in utterance-initial position are chosen to reflect the relative pitch level of a syllable with respect to the next *later* tone-associated syllable. As mentioned earlier, such tones are labeled with colons at their onsets (e.g., :H, :L, or :E) to indicate their status as utterance-initial.

Finally, the fact that RaP entails a consistent relationship between tone labels and perceived pitch means that there are clear correspondences between tone labels and F0 curve characteristics. In the following, we will discuss how RaP labels capture both the gross shape of the F0 curve, as well as the temporal characteristics of the F0 curve, such as the timing of F0 maxima and minima (also known as peaks and valleys). We will break down the discussion by considering particular tone sequences in RaP, starting with high-low and low-high sequences.

3.2 High-low and low-high tonal sequences

In this section, we will discuss the pitch and F0 properties associated with any combination of high and low tones, including H* +L, H+ L*, L* +H, L+ H*, etc. To understand the F0 characteristics associated with a high-low tonal sequence, consider first the properties associated with a (non-utterance-initial) high tone. Recall that a high tone indicates that a particular syllable has a higher pitch than that of the preceding tone. What sort of F0 curve shape will be associated with a high tone? The answer is that a high tone will correspond to a *rise* in F0 (and pitch) from the preceding tone to the high-toned syllable, regardless of whether the immediately preceding tone is on the same syllable or several syllables distant. When the immediately following tone is low, the F0 will subsequently fall. As a result, there will be an F0 maximum (i.e., a peak) in the vicinity of the high-toned syllable, for a high-low tone sequence.

Note that the F0 peak associated with a high tone in a high-low tone sequence may or may not occur on the syllable which is itself marked with a high tone. Rather, it might be delayed very slightly so as to occur in the following syllable's consonantal onset. In spite of such possible F0 "peak delay," the high-toned syllable in a high-low tone sequence should by definition be the perceptually highest-pitched syllable. (If it is not, then a different set of tonal labels is warranted, as discussed later in Section 3.)

A growing body of evidence suggests that listeners differentially sensitive to timing variations in F0 peaks and valleys. (See e.g., House 1990, Dilley 2005.) Some ranges of variation in F0 timing are perceptually salient, while others are not. If desired, an F0 peak or valley which occurs after a H or L tone in a high-low or low-high tone sequence, respectively, can be labeled in the "misc" tier using the symbols ">p" for late peak, or ">v" for late valley, respectively, as shown in <<american>>. In all of the

examples discussed so far, a sequence of a high tone on a target syllable followed by a low tone consistently has described a situation in which there is a locally high pitch on a target syllable, together with a nearby F0 peak.

Next, we can consider the acoustic-phonetic characteristics associated with a low-high tonal sequence, which is the "vertical mirror image" of a high-low tonal sequence. Recall that a low tone phonetically has a lower pitch than that of the preceding tone. Then if the following tone is high, the target low-toned syllable will have the locally lowest pitch. As a result, there will be an F0 valley in the vicinity of the low-toned syllable. This F0 valley may occur on the syllable itself, or it might possibly be delayed slightly so that it occurs just after the target syllable in the following syllable's consonantal onset. In spite of this possible "F0 valley delay," the syllable with the low tone should crucially be the perceptually lowest-pitched syllable in the region.

We have so far established two things. First, a high-low tonal sequence entails the property that the *high*-toned syllable has the locally *highest* pitch in the local speech region, and there will be an F0 peak in its vicinity. Second, a low-high tonal sequence entails that the *low*-toned syllable has the locally *lowest* pitch in the local speech region, and there will be an F0 valley in its vicinity. Next, we will consider how differences in the *temporal alignment* of these pitch and F0 events are captured through RaP labels.

Differences in the timing of pitch events and the alignment of F0 characteristics are captured through the use of starred and unstarred tones. To see this, consider the examples in <<american>> that were presented earlier. In the first example a sequence of a *starred high tone* in the context of a following low captures the fact that *F0 maxima are aligned with metrically prominent syllables*: *mer-* and *ling-*, respectively. That is, the sequence of H* +L on *meri-* and H* +<L on *linguist* each correspond phonetically to a high F0 peak, followed by a fall. Similarly, in the second example a sequence of an *unstarred high tone* in the context of a following low captures the fact that *F0 maxima are aligned with metrically nonprominent syllables*. That is, the sequence of H+ L* on Amer- and +H L* across *ican ling-* once again each correspond to a high F0 peak, followed by a fall. These examples therefore illustrate that whether the high tone in a high-low tonal sequence is starred or unstarred captures the alignment of the high tone with respect to metrically prominent and nonprominent syllables.

We have shown in this section that a high-low tonal sequence entails a locally high pitch on a target high-toned syllable, together with an F0 peak in its vicinity, while a low-high tonal sequence entails a locally low pitch on a target low-toned syllable, together with an F0 valley in its vicinity. We also discussed how a difference in the type of tone (starred or unstarred) captured distinctions in the timing of pitch and F0 events with respect to the syllable sequence. In the following section we will consider the properties associated with tone sequences involving an *equal tone*.

3.3 Equal tones

We now turn to an examination of the acoustic-phonetic characteristics associated with equal tone combinations, including high-equal, equal-high, low-equal and equal-low sequences. We will first consider sequences involving a high tone and an equal tone, followed by sequences involving a low tone and an equal tone.

First, consider the characteristics associated with a sequence of a high tone and a following equal tone. For its part, the high-toned event will have a pitch that is higher than all material between it and the immediately preceding tonally-marked position. Moreover, the following equal-toned syllable will have a pitch that is equal to all material up to and including that syllable. Then a high-equal tone sequence will be characterized by a rise up to the high-toned syllable, followed by a level pitch through the equal-toned syllable. The F0 effect will be a rise up to the high tone, followed by an F0 "corner" marking a change from a F0 rise to a plateau, ending at the equal tone.

To illustrate the perceptual and acoustic properties associated with a sequence of a high and an equal tone, consider <<legumes5>>. In this example, there is a +H tone on the unstressed syllable of *legumes* which occurs in the context of a rightward E* tone. Note that the high tone in this context has a higher pitch than the preceding tone-associated syllable. Moreover, if we imagine an interpolated, smoothed F0 curve, then there is a "corner" associated with the position of the +H tone where the F0 changes from rising to level. The E* tone, for its part, is associated with a pitch which is about the same level as the preceding tonally-marked syllable -*gumes* and all syllables in between. Moreover, if we imagine an interpolated, smoothed F0 curve, then there is another "corner" associated with the position of the E* tone where the F0 changes from level to rising for the following high unstarred tone.

Next, consider the F0 and pitch characteristics associated with a sequence of a low tone and a following equal tone. The low-toned event will have a pitch that is lower than all material up to and including the immediately preceding tonally-marked position. Moreover, the equal tone will have a pitch that is the same level as the preceding tonal event and every syllable in between. The perceptual impression for a low-equal tone sequence is a fall down to the low-tone marked event, followed by a level pitch. In the context of a following equal tone, the low-toned syllable will be realized acoustically in terms of an F0 "corner", which in this case marks a change from a F0 fall to a plateau.

As an illustration of a low-equal tone sequence, consider <<armani7>>. In this case, the !L* tone marks a metrically prominent syllable which has a lower pitch than the preceding syllable. There is a level pitch plateau which extends up until the metrically prominent syllable *mil-* of *millionaire*; this metrically prominent syllable is marked with a E* tone. Thus, a sequence of a low tone followed by an equal tone corresponds to a locally low F0 plateau.

As we have already seen, the distinction between starred and unstarred tones in the RaP system reflects differences in the timing of pitch and F0 properties. In particular, the timing of an F0 "corner" is captured through the difference between E* and E+ (or +E). While <<legumes5>> and <<armani7>> showed that starred equal tones correspond phonetically to F0 corners on metrically prominent syllables, <<i_means2>> shows that unstarred equal tones correspond phonetically to F0 corners on metrically nonprominent syllables. In this utterance, an unstarred E+ tone marks a syllable that is at about the same level as the preceding tonally-marked event, namely, the end of the fall in *I*. Moreover, if we imagine the F0 contour as being smoothed and interpolated, then the E+ tone is associated with a "corner" on a metrically non-prominent syllable which marks a change from level to rising for the following H*. The choice of E+ tone thus reflects the alignment of the "corner" with a metrically nonprominent syllable. In sum, starred and

unstarred equal tones correspond phonetically to F0 “corners” on metrically prominent and nonprominent syllables, respectively.

3.4 Tone sequences of like type

There are a number of restrictions on labeling sequences of like tones in RaP, including high-high, low-low, and equal-equal. This section discusses these restrictions and also presents some examples of contours which are labeled using a sequence of two tones of the same type. In order to convey the rationale for these restrictions, we will first consider the acoustic-phonetic characteristics of high-high and low-low sequences, followed by equal-equal sequences.

What are the acoustic-phonetic characteristics associated with a sequence of two high tones or two low tones? Here, we will consider only cases in non-phrase initial positions. A high-high tone sequence indicates a tone that is higher than preceding tonal material, followed by a tone which is even higher. This will correspond to a sequence of two rises in pitch; there may possibly also be a visible slope change in the F0 contour. An example of an F0 contour which might correspond to a high-high tonal sequence is a rise which starts slow and then rapidly increases. Similarly, a low-low tone sequence indicates a tone that is lower than preceding tonal material, followed by a tone which is lower still. This corresponds to a sequence of two falls in pitch. Again, there may be a change in slope of the F0 contour.

To illustrate these characteristics, consider the example in <<mama_lemm>>, which illustrates both high-high and low-low tone sequences. Here, the end of the phrase shows a locally high pitch on a metrically weak syllable, (*Lem*) followed by a small drop in pitch to a stressed syllable, followed again by a steep drop in pitch. This is labeled as H+ L* L. Note that there is a change in overall slope in the vicinity of the L* tone. In addition, there is a portion of the beginning of the contour which involves a shallow rise to a relatively higher pitch, followed by a steeper rise to a considerably higher pitch. This initial portion of the contour is labeled L H* H+. Note the change in slope in the vicinity of the H* tone.

Another restriction on labeling high-high and low-low tone sequences is that the first tone in the sequence must be a starred tone. As a result, sequences of two high tones must take the form H* H… Examples of permissible sequences of two high tones include H* H+, H* +H, H* H, H* H*, etc. Similarly, sequences of two low tones must take the form L* L… Examples of permissible sequences of two low tones include L* L+, L* +L, L* L, L* L*, etc. This restriction is based on the idea that the changes in F0 slope which mark transitions from one high tone to another or from one low tone to another are more likely to be detected on metrically prominent syllables than in other positions.

Note that RaP permits no more than two high tones or two low tones to be labeled in adjacent positions in a tonal sequence. This restriction on tone labels is related to the phonetic characteristics associated with identical tone sequences that are described above. In particular, there is no salient F0 characteristic which marks the transition from a high tone to another high tone, or from a low tone to another low tone. As a consequence, it will sometimes be difficult to distinguish whether two high tones or two low tones should be labeled, or just one. The RaP system limits further such uncertainties by not permitting multiple high tones or multiple low tones to be labeled in a row.

In the preceding discussion, we have highlighted the significance of a slope change as the expected phonetic correlate of a high-high or low-low tone sequence. What are the implications of these observed F0 correlates for sequences of equal tones? By extension of previous arguments, two adjacent equal tones are expected to give rise to speech with flat pitch and level (monotone) F0. In other words, no F0 slope change is expected to occur at any time at the juncture of two equal tones. As a result, there is no way of distinguishing a monotone F0 which might arise from a single equal tone from a monotone F0 which arises due to two or more equal tones. The RaP system eliminates this ambiguity by disallowing the labeling of equal tones in adjacent positions: every equal tone must be followed by a high tone or a low tone.

Note that we have so far considered restrictions on sequences of identical tones which are not in phrase-initial position. An important point is that in phrase-initial position, the maximum number of like tones which is allowed in sequential positions is increased by one relative to non-phrase-initial position. As a result, a maximum of three adjacent high tones or low tones may occur in phrase-initial position; the first of these three adjacent identical tone types must be the initial tone in the new phrase. Moreover, a maximum of two adjacent equal tones may occur in phrase-initial position.

It will sometimes be challenging to determine whether a sequence of two high tones or two low tones should be labeled, or whether a single high or low tone should be selected, since the phonetic characteristics associated with two like tones versus one tone are similar. There are two principles which can be applied to aid in making labeling decisions. The first is a general principle which applies to many different labeling situations. It is *when in doubt, a simpler transcription should be preferred to a more complex one*. As a result, a single low tone or high tone should be preferred to a sequence of two low tones or two high tones. The second principle relates to the degree of prominence across a target syllable. Recall that whenever a sequence of two high tones or two low tones is labeled, the first of those tones must be a starred tone. If the syllable which is being considered for the starred tone in the two-tone sequence is very prominent and is labeled with "X", then the two-tone interpretation should be preferred. Under this interpretation, the extra-prominent syllable will be labeled with a starred tone, while the following syllable will have a tone of like type, giving H* +H, L* +L, etc. In contrast, when the syllable being considered for a starred tone in the two-tone sequence is only moderately prominent and is assigned an "x", then the single-tone interpretation should be preferred.

To see how these conventions can be applied in practice, consider the example in <<handicap>>. Here, there is some ambiguity as to whether *first after it* should be labeled with two like tones, H* L* L+, with a L* tone on *af-* of *after*, or simply as H* L+, with no additional low tone on *after*. The simpler H* L+ sequence is sufficient to describe the contour, and the rule of thumb is that a simpler transcription is preferred whenever there is any doubt regarding label selection. Moreover, the syllable *af-* of *after* is indicated to have only moderate prominence, as indicated by the small "x" in the "rhythm" tier. This provides additional impetus for selecting H* L+ instead of H* L* L+ to describe the pitch contour on *first after it*.

Another example of how these conventions can be applied comes from <<mathematics>>. In this example, the phrase *some of the mathe-* involves an extended fall which could be labeled either as H* L* L+ or as H* L+. In this case again, the

simpler labeling will suffice to describe the contour shape. Moreover, the syllable *math-* in *mathematics* only has moderate prominence, providing additional justification for selecting H* L+ as the labeling in this case. Had *math-* been labeled as having a very metrically prominent beat ("X"), then the labeling H* L* L+ could have been selected instead for this phrase. This concludes the discussion of labeling tones of the same type in sequence.

3.5 Tone label uncertainty

Several types of tone label uncertainty can arise in the process of evaluating a speech utterance. The first type of ambiguity we will discuss concerns whether a tone should be labeled on a given syllable or not. For example, there may be uncertainty in whether an unstarred tone should be labeled on a particular unstressed syllable. This issue is exemplified in <<armani9>>. In this utterance, *mil-* is clearly low in pitch. However, it is unclear perceptually whether the preceding syllable is at the same level in pitch as *mil-*, or whether *mil-* continues the fall through the phrase. If the preceding syllable is identified as being clearly at the same level in pitch as *mil-*, then an E+ tone should be labeled on *the*. On the other hand, if *mil-* continues the preceding fall, then no E+ tone should be labeled on *the*. Listening multiple times to this phrase in context does not seem to clearly resolve the percept one way or another. (Note that it is important not to attempt to decide by playing just one or two syllables in isolation. It is always necessary to listen to at least several syllables of context.) The solution to labeling tones in <<armani9>> is to label "+?" on the word *the* to indicate uncertainty regarding whether an unstarred tone is warranted. Likewise, there are other instances in which there will be uncertainty regarding whether a starred tone is warranted. In such cases, the symbol "*?" can be indicated.

A second type of ambiguity that can arise concerns the relative pitch of a syllable with respect to other syllables in sequence. Often the choice will be between two relations: higher vs. equal, or lower vs. equal. When it is possible to narrow the tone label down to Sometimes it will not be possible to determine with any certainty whether a particular position in the speech stream has a pitch that is higher, lower, or equal than the preceding material. If this happens, there are several labels that can be used to indicate uncertainty. If a labeler suspects that a particular syllable either has a higher pitch or an equal pitch relative to some preceding tonal referent, then the label "H?*" should be used for a prominent syllable, while the labels "H?+", "+H?" or "H?" should be used for a nonprominent syllable. Likewise, if a labeler feels that a particular syllable either has a lower pitch or an equal pitch relative to some preceding tonal referent, then the label "L?*" should be used for a prominent syllable, while the labels "L?+", "+L?" or "L?" should be used for a nonprominent syllable.

A third type of ambiguity that may occasionally arise is the following. A labeler may be certain that a particular syllable should be marked with a tone but uncertain about whether that syllable is higher than, lower than, or equal to the contextual referent. In such a situation, an "X*" or "X+" label should be indicated on the "tones" tier in place of a regular high, low, or equal starred or unstarred tone label. One case in which labels like X* and X+ are often useful is when the voice is creaky and thereby disrupts the pitch, as

in <<smart>>.[5] In this example, the pitch across *young thing I 'spose* has unclear pitch due to the intermittent modal voicing. Thus, a X* is indicated at the right edge of the creaky region.

Note, though, that a creaky voice is usually a low voice. Thus, L*, L+, or +L can often be used to describe a drop in pitch to a low, creaky voice. Similarly, creaky voice is often maintained rather consistently over some portion of a speech utterance. When creaky voice is rather steady without intermittent returns to modal voicing, then an E* or E+ should be labeled at the right edge of the creaky region. Finally, we note that whenever possible, low and equal tones are preferred over X* or X+ labels.

3.6 Some additional conventions for tone labeling

Now that we have discussed some of the perceptual and acoustic-phonetic properties associated with tones in the RaP system, we are in a position to consider tonal labeling conventions in more depth. In the following, we will focus on conventions for labeling tones in phrase-medial position. In particular, we will address when and how to label *unstarred* tones in phrase-medial and phrase-final positions.

Recall first that a phrase-medial unstarred tone must be labeled in a position that is *next to* a starred tone. In most cases, the unstarred tone and starred tone will be on different syllables. When this happens, the unstarred tone will be labeled on a weak syllable that immediately precedes or follows a starred, prominent syllable. However, in some cases the unstarred tone may be labeled on the *same* syllable as a starred tone; the distinction depends on the timing of the unstarred tonal event with respect to syllable boundaries. As illustrated in previous examples, the "+" symbol is used in conjunction with phrase-medial unstarred tones to indicate the relative position of an unstarred tone relative to a starred tone. In particular, the "+" is written to the left of the unstarred tone (giving +H, +L, or +E) whenever the preceding metrically prominent position is starred. Moreover, the "+" is written to the right of the unstarred tone (giving H+, L+, or E+) whenever the following metrically prominent position is starred. When both the preceding and following syllables have a starred tone, the "+" is written to the *right* of an unstarred tone.

To illustrate these conventions, consider again the examples in <<maria>>. In the first utterance in this example, the unstarred L (and :L) tones occur on the metrically weak syllables *It's* and *Ma-* which precede starred syllables. As a result, these tones are notated with a "+" on their right, to indicate that the starred tone occurs on the following syllable. Similarly, in the second utterance, unstarred H and (and :H) tones are on metrically weak *It's* and *Ma-* and precede starred syllables. Crucially, note in the third utterance that the location of the medial syllable with the pitch peak has now switched to *-ther* in *mother*, as compared with its place on *Ma-* in the second example. As the result, the medial pitch peak now is adjacent to a starred syllable to its *left*, so it is labeled with a "+" to its left, giving +H.

Another convention in RaP relates to the annotation of tones which are produced as very high or very low in the speaker's pitch range. Such tones should be assigned the

[5] :X* or :X+ can also be used at the left edge of an utterance to denote a syllable produced entirely with creaky voice.

symbols “>” or “<” to indicate that the tone is very high in pitch or very low in pitch, respectively, for the speaker’s voice. Thus, a speech utterance which ends in a very low fall would be labeled with “+<L” or “<L”, while a speech utterance which ends in a very high rise would be labeled with “+>H” or “>H”. (Note that conventions on the use of “+” for phrase-final tones will be discussed in an upcoming section.) Similarly, phrase-initial tones or phrase-medial pitch accented tones which are very high or low in pitch should be assigned these markers. In general, “<” and “>” will not be used in conjunction with equal tones, since in most cases, such specifications are preceded by a rapid rise or fall.[6] The use of these markers is illustrated in the utterance <<bananas>>.

Labeling practice. Tone selection and conventions for tone labeling.
(Ignore interval sizes and phrase-final tones for the following examples and focus on simply selecting the right tone and other diacritics.)
<<armani11>>
<<armani1>>
<<ma_lemm>>
<<money>>
<<how_long>>
<<phone>>
<<how_are_you>>
<<wait>>
<<slate>>
<<pro_ball>>
<<insects2>>

3.7 Pitch range and interval size

It is well-known that speakers use pitch range in expressive and meaningful ways. For example, focused elements in sentences often have an expanded pitch range. The local expansion in pitch range aids in directing the listener’s attention to the important information in the sentence. RaP captures this meaningful variation by providing conventions for annotating local pitch range expansion and compression.

The use of a locally expanded pitch range to highlight important information is illustrated in <<blue_ties1>> and <<blue_ties2>>. Note that in each of these examples, the contrastively focused item (*blue* or *ties*) involves a larger pitch excursion than the rest of the utterance. The examples in <<blue_ties1>> and <<blue_ties2>> show how local

[6] An exception is that “>” and “<” symbols should be used in conjunction with level plateaus which are high or low in the pitch range, respectively, which are not preceded by a rise or a fall. For example, a high, level plateau not immediately preceded by a rise would be labeled with “>:E +>E” or “>:E >E”, while a low, level plateau not preceded by a fall would be labeled with “<:E +<E” or “<:E <E”.

changes in the sizes of pitch excursions are used in important ways in spoken communication.

These meaningful differences in pitch range are captured in the RaP system by the use (or non-use) of the "!" diacritic. When the "!" diacritic is labeled before a starred or unstarred tone, it indicates that the tone exhibits locally *reduced* pitch range. To put it another way, the "!" diacritic indicates that the tone participates in a small pitch interval. On the other hand, the lack of a "!" diacritic before a tone indicates that the tone exhibits a relatively *expanded* pitch range, or that it participates in a comparatively larger pitch interval.

The decision of when to use the "!" diacritic depends on the *perceived degree of pitch change* at a particular tonal position relative to the preceding pitch context. As mentioned above, the "!" diacritic is used when the degree of change in pitch is *small*. In contrast, the "!" diacritic is left off when the degree of pitch change is *large*. In the following, we will consider some criteria for what constitutes a *large* vs. a *small* pitch change and when to use the "!" diacritic.

There are several characteristics that are used to determine if the pitch range is locally expanded or compressed. The first is the size of the pitch interval perceived between the target tone and the preceding tonal context. RaP defines the basic distinction between "small" and "large" pitch intervals in musical terms. In particular, a "small" interval is defined as one which is 1-2 semitones, while a "large" interval is one which is 3 semitones or greater. The speech utterances in <<1_semitone>> and <<2_semitone>> illustrate pitch changes of 1 and 2 semitones, respectively. Each of these changes constitute "small" pitch intervals in RaP and would warrant the use of the "!" diacritic on the appropriate tones. In contrast, examples of "large" pitch intervals are illustrated in <<3_semitone>>, <<4_semitone>>, and <<5_semitone>>. These examples illustrate pitch intervals of three, four, and five semitones, respectively.

A second criterion for whether a tone should be marked with the "!" diacritic is the shape of the pitch trajectory leading up to the target tone itself. If the pitch change from the last tone to the target tone is *gradual* and *smooth*, the pitch interval is classified as small and the "!" diacritic is indicated. In contrast, if the pitch changes *rapidly*, or if there is a *jump up or down* to the target tone, the pitch interval is classified as large and the "!" diacritic is left off. For example, if there is a local jump in pitch marking a departure from the overall pitch trajectory, the interval is classified as large.

To illustrate how these criteria are applied in practice, consider the example in <<armani4>>. The beginning of this example is marked by large pitch excursions on the focused element, *Armani*. The pitch range on *knew the millionaire* is locally reduced, and there is a very gradual fall from *knew* to *–naire*. As a result, the phrase-final low tone on *-naire* is assigned the "!" diacritic.

Another example illustrating the difference between "small" and "large" intervals comes from <<i_believe>>. In this example, the pitch gradually rises from the initial tone in the utterance up through the tone on *-lieve*, consistent with a !H* tone. Similarly, the pitch gradually falls up through *box-* of *boxcars*, consistent with a !L* tone.

Next, consider the example in <<legumes2>>. In the initial part of the file, there is a very compressed pitch range with small pitch excursions between syllables. As a result, all tones are assigned "!" diacritics. The nuclear accent on *vit-* exhibits an

expanded pitch range and hence is not marked with a "!" diacritic. Similarly, the phrase-final tones fail to show a reduced pitch range and hence are not labeled as small intervals.

A radio broadcast example illustrating pitch range distinctions comes from <<musicians>>. Throughout most of this example, the speech involves small pitch intervals, which are labeled with "!". Note, however, that large pitch intervals are employed on information-bearing words and phrases: *musicians*, *car engines*, *should sound like*.

Another example comes from <<asylum>>. Consider the initial portion of the speech, which is labeled as :L H*. Note that across *but they* there is a small rise of about 2 semitones, but then the pitch jumps up for *can't*. Such a pitch jump is an indication that a small interval should not be labeled. On *asy-* of *asylum* the pitch interval is small, however. Next, consider the word *federal*. Here, the F0 falls fairly slowly, but the auditory impression is of a pitch jump from *fed-* to *-ral* (3 semitones). Thus, this is not treated as a small interval.

Two points relating to notation are worth mentioning. First, note that "!" tones are only marked on low or high tones, never on equal tones. This is because equal tones by definition involve a *compressed* pitch range, as indicated by flat pitch. Because the "!" diacritic would be redundant when used in conjunction with an equal tone, it is left off. The second point relates to tones which are initial in the utterance and which are assigned the ":" diacritic. Recall that the choice of high, low or equal for such tones depends on the relative pitch of the initial syllable with respect to the *following* tone, as opposed to the preceding tone. As a result, the use of the "!" diacritic in conjunction with such tones depends on the perceived pitch interval with respect to the following tone as well. In particular, such tones are labeled with "!" whenever the following tone is itself labeled with "!".

There will sometimes be ambiguity in whether an interval is "large" (three semitones or greater) or "small" (two semitones or less). In this case, a "?" is placed *before* the tone label to indicate this uncertainty. In general, several phonetic properties of syllables probably contribute to a propensity for uncertainty in the size of the interval in most cases. First, a syllable which is short will leave little time for the listener to develop a clear pitch percept of a syllable against which to judge the pitch of another syllable. Second, when the target syllable is followed by a sonorant segment, such that there is a smooth rise or fall into the following syllable, then the lack of spectral discontinuity will cause the pitch of the target syllable to be elusive perceptually.

The example in <<impression>> illustrates such an ambiguity. This example exhibits an expanded pitch range through most of its extent, so that the "!" diacritic is left off the associated tones. There is a small portion of speech toward the end of the file which shows some local reduction in pitch range. It is straightforward to label *-ble chunk* in *sizeable chunk* as L+ !H*; the pitch interval is approximately two semitones in musical terms. However, it is hard to say with certainty whether the pitch interval from *chunk* to *of* in *chunk of money* is closer to two semitones or to three. As a result, the uncertainty label "?" is applied.

Labeling practice: Pitch range and interval size.

<<armani6>>
<<armani9>>
<<curve>>
<<friend>>
<<elmira>>
<<hyannis>>
<<insects1>>
<<honda>>
<<bank>>

3.8 "False" pitch accents

In most cases, pitch excursions which occur on metrically prominent, stressed syllables make those syllables sound accented and act to highlight them in a discourse. Pitch excursions of this sort are referred to as *pitch accents*. However, there are occasionally situations in which there is a pitch excursion on a metrically prominent syllable, where that syllable *fails* to be a pitch accent. The present section deals with the labeling conventions for these "false pitch accents".

We have already established a set of conventions for indicating pitch excursions on metrically prominent or stressed syllables. In particular, "x" or "X" is used on the "rhythm" tier to indicate metrical prominence, while starred tones (H*, L*, or E*) are used on the "tones" tier to indicate pitch excursions on these syllables. There are two distinct situations in which a pitch excursion can occur on a metrically prominent syllable, such that the result is *not* a pitch accent. We discuss each case below. These false pitch accents are notationally distinguished in the RaP system from actual pitch accents by enclosing the asterisk of a starred tone in square brackets: "[*]".

The first type of situation in which a false pitch accent can arise is termed *backgrounding*. Backgrounding occurs in the context of alternating high and low pitches on metrically prominent, accentable syllables. In such a circumstance, low, starred tones will often fail to sound accented and hence are "backgrounded". The name for this phenomenon was chosen to reflect concepts from Gestalt psychology in which certain elements in an auditory or visual scene receive less attentional focus and hence are backgrounded relative to other elements in the scene. Our working hypothesis is that backgrounding *only* occurs in American English on low, metrically prominent syllables in the context of alternating high, prominent syllables, as illustrated in the following examples.

First, consider the example in <<armani2>>. In this speech utterance, the syllables *ma-*, *knew*, *mil-* and *naire* each sound metrically prominent. Moreover, these syllables have locally high, low, high, and low pitches, respectively. Here, only the high-toned syllables *ma-* and *mil-* sound accented, while low-toned *knew* and *naire* do not sound accented. Both *knew* and *naire* are backgrounded; *knew* is labeled with a starred tone enclosed in square brackets to indicate that it is a false pitch accent. (Note that a separate set of conventions apply to labeling tones on phrase-final syllables, such as *naire*, as discussed in the following section.) Finally, note that syllables which are

backgrounded will always occur in conjunction with a metrically prominent syllable labeled as "x", never as "X".

Another example of backgrounding is illustrated in <<armani3>>. In this example, there is again a sequence of high and low alternating tones on metrically prominent syllables: *ma-*, *knew*, *mil-* and *naire*. Here again, only the high-toned syllables sound like accents; the low tones don't seem to add any extra prominence to *knew* and *naire* in this context. Thus, the asterisk of the starred low tone on *knew* is enclosed in square brackets. (Note that the labels on the final low-toned syllable *naire* are governed by conventions for labeling tones in phrase-final position, as discussed in the following section.)

A third example of backgrounding comes from <<tape_machine>>. Once again, the initial part of this utterance shows a sequence of metrically prominent syllables: *tape*, *-chine*, *-cords*, and *well*. Moreover, these utterances show a repeating up-and-down pattern of high and low tones, where the low-toned syllables do not sound accented. These syllables are consequently labeled with starred tones enclosed in square brackets.

A final example illustrates that backgrounding can occur even when there is just a single high, prominent syllable, followed by a low. In <<armani4>>, there is a high pitch accent on *ma-*, and the following metrically prominent syllable *know* is low. The pitch change on *know* fails to make it sound accented. Hence, this low tone is assigned a bracketed star.

Earlier it was stated that there are two types of situations in which false pitch accents arise. The second is termed a *prominence mismatch*. This occurs when a metrically prominent syllable exhibits a pitch excursion that cannot be interpreted as a pitch accent due to the rules of lexical and phrasal prominence in English. In other words, the listener cannot "hear" the syllable as a pitch accent because doing so is incompatible with listener's implicit knowledge of prominence placement.

One common situation where a prominence mismatch can arise involves a complex pitch pattern near the end of a phrase. In particular, when the pitch pattern shows a LHLH pattern, the second L can often correspond to a prominence mismatch. For example, consider <<abercrombie_LH>>. In this example, there is a LHLH at the end of a phrase; the second L occurs on the secondary stressed syllable *crom-*. It is well-known that pitch accents are not allowed on "postnuclear" secondary stressed syllables, i.e., syllables with secondary stress which occur after the main prominence for the phrase. Native English speakers know this implicitly, and will interpret pitch excursions in such positions as non-accents. As a result, *crom-* cannot be a pitch accent, and the associated low tone is assigned the symbol "[*]."

A complex LHLH pitch pattern can also arise across a shorter word so as to induce a prominence mismatch. For example, consider <<anna_incredulous>>. In this example, the entire LHLH pattern occurs across the word *Anna*, with the second L tone occurring on the unstressed syllable *na*. The result is that *-na* sounds metrically prominent; because of its lexically unstressed status it is accordingly assigned a [x] in the "rhythm" tier. Moreover, the listener "knows" that unstressed *na* is not a possible location for a pitch accent in English. The syllable *na* therefore corresponds to an instance of prominence mismatch. As a result, the low tone on *na* is assigned "[*]" on the "tones" tier.

The example in <<heavy-rain>> illustrates another case of a prominence mismatch. Here, the unstressed unreduced syllables *-ble* of *possible* and *-ty* of *seventy* sound metrically prominent in context. However, since these syllables are unstressed unreduced and not secondary or primary stressed syllables, they are labeled as "[x]" rather than "x". Note that the tones which occur on these two unstressed unreduced syllables seem to inherit the prominence of these quasi-prominent syllables. As a result, bracketed starred tones are assigned to the syllables.

Finally, it is important to note that the "[*]" symbol is used only in conjunction with pitch excursions on metrically prominent syllables exhibiting characteristics of backgrounding or prominence mismatch, as described above. In particular the "[*]" label should not be used as a marker of uncertainty regarding the accentual status of a given tone, nor should it be used as a marker of a lesser degree of prominence in general. Rather, its usage is restricted *only* to the cases described above.

3.9 Labeling phrase-final tones

This section deals broadly with the issue of how to label tonal patterns in the vicinity of major and minor phrase boundaries. We will first consider conventions for labeling tones at the ends of phrases, followed by conventions for labeling tones at the beginnings of phrases.

In general, the RaP system assumes a high degree of independence between rhythm labels and tonal labels. In particular, labeling a phrasal boundary in the "rhythm" tier does not require a labeler to annotate a phrasal tone in the "tones" tier. In this respect, RaP contrasts with ToBI, which instead requires that a phrase boundary tone be marked at every point of perceived disjuncture, even in the absence of a local pitch change. The fact that RaP does not require a tone to be labeled at every phrasal boundary is consistent with the fact that perception of greater disjuncture between words or syllables can arise through a variety of different phonetic dimensions, not just tonal movement. For example, a sense of greater disjuncture between words can be brought about by durational lengthening or voice quality changes, with or without accompanying tonal variation. Thus, RaP permits labelers the option of labeling tones at perceived phrasal boundaries when the tonal evidence warrants such a label, rather than forcing labelers to assign a tone at every phrasal boundary.

As a result of the independence between the "tones" and "rhythm" tiers, there are only a few circumstances in which RaP *requires* that a tone be labeled. One such instance concerns syllables which are at utterance boundaries. The convention adopted is this: *syllables which are at the beginning or end of an utterance must be labeled with at least one tone*. Similarly, there is a convention related to phrase-initial and phrase-final syllables. Under the simplifying assumption that acoustic silence induces the listener to hear a phrasal boundary, the convention is that *a syllable which is preceded or followed by a pause (and hence is at the edge of a phrase) must be assigned at least one tone.*[7] It might be observed that the former convention is entailed in the second, but we distinguish these two conventions because of further distinctions between phrasal and utterance labeling conventions. Moreover, we will return to the issue of what type(s) of tone(s) should be labeled when there is a following pause shortly.

[7] The pause duration should be at least 100 msec.

The converse of the conventions given above is that when there is *no* pause following the phrase-final syllable (or one which is less than 100 msec), RaP does not *require* a labeler to indicate *any* tones on that syllable. Instead, decisions about whether tones are present are made based on the tonal evidence across the syllable in question. In particular, if there is smooth interpolation across the phrase-final syllable, so that the pitch contour across the phrase-final syllable smoothly falls, smoothly rises, or remains flat, then no tone should be indicated. On the other hand, if there is a local change in the direction of the pitch contour, then one or more tones should be labeled.

Suppose that we determine that at least one tone should be labeled at the right edge of a phrase. As mentioned already, at least one tone should be labeled in either of two situations: (1) when there is a pause following the phrase final syllable, or (2) when the speech is continuous (i.e., there is no pause) but there is a local change in the direction of the pitch contour on the phrase-final syllable. How then should these phrase-final tones be labeled? In the following paragraphs we discuss several relevant conventions.

First, we will consider cases in which only a *single* tone is warranted. A single high, low, or equal tone should be used whenever a phrase-final syllable marks the ending point of a smoothly rising, smoothly falling, or level contour. (We note that simple rising and falling intonation contours typically correspond in English to questions and statements, respectively.)

The metrical prominence status of a phrase-final syllable will affect the type of tone selected at the end of a phrase. We'll first consider cases in which the final syllable is metrically nonprominent. To illustrate simple rising and falling phrase-final tonal movements warranting a single tone, consider the utterances in <<anna1>>. The last syllable in each utterance is metrically nonprominent. In the example on the left, there is a simple falling intonation pattern at the end of the phrase. This is labeled with a *low unstarred* tone; additionally, the "<" symbol is used to indicate that this tone occurs in the lowest part of the speaker's range. In the example on the right, there is a simple rising intonation pattern at the phrase boundary. This rise is labeled with a *high unstarred* tone, where the ">" symbol indicates that the tone occurs in the highest part of the speaker's range. Finally, note that conventions described earlier for marking the "+" symbol are adhered to here. In particular, unstarred tones are marked with a "+" symbol indicating the relative position of the starred tone on a syllable immediately to their left.

Note that in both utterances in <<anna1>>, no special diacritics are used to indicate the phrase-final tone; diacritics like "+", "<", and ">" are not confined to phrase-final position. The status of the tone as phrase-final is recoverable from its position with respect to any phrasal boundary which is labeled in the rhythm tier. This convention of RaP contrasts with the ToBI system, which assigns redundant special diacritics to phrase-final tones.

Next, consider how to label a simple level contour at the end of a phrase across a nonprominent syllable, as illustrated in <<money>>. Here, the level contour extends across the latter half of the phrase. A simple "E" tone is labeled at the end; no "+" is indicated, because there is a starred tone neither on the same syllable, nor on an adjacent syllable.

In <<anna1>> and <<money>>, the phrase-final syllable was metrically *nonprominent*. Next, we will consider how to label a single tone on a phrase-final syllable

which is metrically *prominent*. In such a case, the labeler has a choice regarding whether to label an unstarred tone, a starred tone, or a bracketed starred tone, or a combination of these tones. The choice of tone type will depend largely on the shape of the contour across the final syllable, along with the degree of perceived prominence on the final syllable. We will first consider when an *unstarred* tone should be labeled on a phrase-final, metrically prominent syllable, and then we will consider when a *starred* tone should be labeled.

If the pitch contour rises or falls smoothly to the very end of a phrase-final metrically-prominent syllable and the pitch movement does not seem to lend extra prominence, an *unstarred* tone should be labeled. For example, in <<armani9>> the final rise starts on *mil-* and continues to the last voiced portion of *-naire*. Crucially, the rise continues *all the way through* the final syllable; moreover, the last pitch accent is *mil-*, not *-naire*. Under these conditions, an *unstarred* high tone should be labeled.

In contrast, *if the pitch contour levels off at the end* of a phrase-final, metrically-prominent syllable, a *starred* tone should be labeled (or bracketed starred tone, following criteria outlined earlier). For example, in <<armani12>> there is a rise that starts on *mon-* , but the last syllable shows a decided leveling off of the pitch. In contrast to <<armani9>>, the rise in this example does not continue to the very end of the phrase. Instead, the pitch across *suit* sounds generally high but crucially does not seem to rise throughout its extent. The auditory impression is of a clear, level pitch at the end of the phrase. When the pitch contour at the end of a phrase on a metrically prominent syllable levels off in this way, so as not to rise continuously to the very end of the phrase-final syllable, a *starred* high tone should be labeled.

The fact that the shape of a contour across the final syllable affects labeling choices is further illustrated in <<armani5>> and <<armani2>>. In <<armani5>>, the pitch falls smoothly to the end of phrase. In this case, an unstarred tone is selected to reflect the fact that the pitch falls to the very end of the phrase. In <<armani2>> the end of the phrase on *-naire* shows a simple fall which levels off at the end. A starred tone would ordinarily be warranted, but the overall pattern of alternating high-low pitches on successive metrically prominent syllables leads to the choice of a bracketed starred tone instead.

It will not always be easy to decide whether the pitch rises or falls smoothly to the end of the phrase, or whether it levels off. *If there is any uncertainty about the type of tone to be selected for a phrase-final metrically prominent syllable*, *label a starred tone (or bracketed starred tone).*

An illustration of conventions for labeling simple rises and falls on both prominent and nonprominent syllables at the ends of phrases comes from <<hands_tied>>. First, consider the minor phrase boundary on *says* which is coupled with a simple fall. The fall is captured by an unstarred low tone. There is no following pause, and the following syllable has a starred tone, so a "+" is indicated, giving L+. Next, there is a major phrase boundary after *owners*, which shows a simple rise on the nonprominent syllable *-ners*. This is captured using a high unstarred tone. Moreover, since there is a starred tone on a preceding syllable (but not a following syllable), a "+" is indicated to the left of the high tone, giving +H. Finally, there is a steady fall at the end of the utterance on the metrically prominent syllable *tied*. The fall continues through the entire syllable and an audible drop in pitch is heard. As a result, an unstarred low tone is

assigned. The fact that there is a starred tone to the left and that the pitch drops to the bottom of the range leads to the addition of "+" and "<" diacritics to this tone.

So far we have been considering conventions for labeling a simple rise, fall, or level contour, all of which require only one tonal label. What if the tonal movement is more complex? For example, what if the pitch contour following the last pitch accent in the phrase falls and then rises? In such a case, two tones would be required. In the following we will consider conventions for labeling phrase-final pitch contours using two tones.

First, consider the case of two tones being required on a phrase-final *nonprominent* syllable. The two tones will, of course, be unstarred. Rules for assigning "+" are the same as in phrase-medial position. That is, if the first unstarred tone is adjacent to a syllable with a starred tone, a "+" should be indicated to the left of the tone. The second unstarred tone should be assigned a "+" only when (1) there is no following pause, and (2) the initial syllable in the following phrase carries a starred tone.

Next, consider the case of two tones being required on a phrase-final prominent syllable. In this case, the first tone will always be starred (or a bracketed star), while the second tone is unstarred. The conventions for the use of "+" with the final unstarred tone are consistent with other conventions described elsewhere in the manual. If the following syllable has a starred tone and there is no following pause, a rightward-aligning "+" should be used. By contrast, if the following syllable lacks a starred tone, a leftward-aligning "+" should be used. In the next section we will consider conventions for labeling phrase-initial tones.

3.10 Conventions for labeling phrase-initial tones

Now that we have considered how to annotate phrase-final tonal information, we can consider how to annotate phrase-*initial* tonal information. In particular, this section addresses two issues. First, we will discuss *when* a tone should be labeled on a phrase-initial syllable, as well as when a tone should not be labeled. Second, we will discuss *what kind* of tone should be labeled on a phrase-initial syllable.

As mentioned already, phrasal boundaries are associated with a variety of phonetic characteristics. Some are marked by local tonal variation, while others are not. Similarly, some are demarcated by pauses, while others are not. In some cases, the end of one phrase blends gradually into the start of the next, with no pause or other interruption. Such cases raise the possibility that there is simply phonetic interpolation between a phrase-final tone and some later tone in the sequence. This interpolation may even occur across a phrase-initial syllable. *Whenever there is smooth interpolation across a phrase-initial syllable, that syllable should not be labeled with a tone.* For example, suppose a phrase-initial syllable occurs in the middle of a rising, falling, or level stretch of pitch, where there is no pause or other interruption between the end of the previous phrase and the start of the new phrase. If the rise, fall, or level contour began before the phrase-initial syllable and ends after the phrase-initial syllable, this constitutes interpolation across the phrase-initial syllable itself. In such a case, no tone label should be indicated on the phrase-initial syllable.

On the other hand, *there are two circumstances in which a tone label must be indicated on a phrase-initial syllable:* (1) *when a phrase-initial syllable marks a local*

change in pitch, i.e. there is no interpolation across this position, and (2) *when the phrase-initial syllable is preceded by a pause (of greater than 100 msec).*

There is another principle which guides the choice of label selection in phrase-initial, and phrase-final position, as follows. *A major phrase consists minimally of two tones – one on the first syllable in the phrase, and one on the last syllable in the phrase.*

When selecting a tone label for a phrase-initial syllable, should a tone label be selected relative to the tone at the end of the preceding phrase, or relative to the following tone? The answer is that if the phrase-initial syllable is preceded by a *long pause* (of at least 1 second in length), then it counts as "utterance initial" and conventions for labeling utterance-initial tones should be applied. In particular, the ":" symbol should be used in conjunction with the tone label, and the tone's identity will reflect the relative pitch level of the phrase-initial syllable with respect to the following tonal marker. On the other hand, if the phrase-initial syllable is preceded by less than a second of silence, the tone label is selected to reflect the relative pitch level of the phrase-initial syllable with respect to the syllable at the end of the preceding phrase.

To see how the conventions apply for labeling phrase-initial and phrase-final tones, consider the case of <<avon>>. First, the syllable *from* is both phrase-initial and phrase-final. This syllable is metrically prominent, but it does not continue an earlier rise, fall, or level contour. Thus, it is assigned a starred tone, along with the ":" diacritic, since the tone is also utterance-initial. Note that this syllable also exhibits a slight fall, which must be captured through the labeling of an additional unstarred low tone. This tone is assigned a "+" diacritic to its left, since there is a preceding starred tone on the same syllable, yielding "+!L". Next, there are major phrase boundaries after *Avon* and *Corning*. Neither item has a phrase-final syllable which is metrically prominent, so these syllables are labeled with unstarred tones. In each case, there is a preceding starred tone, such that a "+" is indicated to the left of each unstarred tone. Next, consider the phrase-initial syllable *to* in the phrase *Avon to Corning.* This syllable does not exhibit a local change in pitch; rather, there is a smooth, rising interpolation from the preceding syllable, *-von* of *Avon*, up through the locally high pitched syllable *Cor-* of *Corning*. As a result, no phrase-initial tone is labeled. Similarly, the phrase-initial syllable *it* in *it'll* also does not mark a local change in pitch, so it, too, does not receive a phrase-initial tone marker. Similarly, the phrase-initial syllable *to* in *hours to get* seems to be situated in the middle of a flat-pitched region, so it also fails to get a tone. Finally, the phrase-final syllable *there* in the phrase *get there* is not metrically prominent, so it gets an unstarred tone. Because the preceding syllable has a starred tone, the "+" diacritic is assigned to the tone's left. Finally, note that the syllable falls to the bottom of the speaker's pitch range, so the "<" diacritic is additionally assigned to this tone.

Another example of labeling phrase-initial and phrase-final tones is <<graft>>. We will consider the choice of tonal labels at each of the indicated phrasal boundaries in turn. For the boundaries after *oh the* and *there's a*, no metrical prominence occurs on the phrase-final syllable. As a result, an unstarred tone is labeled on each syllable. These unstarred tones are assigned a "+" to their left, giving +!H and +!L, due to the fact that there is a following pause of greater than 100 ms. Next, there are major phrasal boundaries after *joke*, *the*, and *graft*, and each of these syllables is metrically prominent. The word *joke* marks a simple rise from the preceding syllable, so a single high starred tone is indicated on that syllable. The word *the* counts as the only syllable in a major

phrase, and it must carry two tones. A combination of a starred and an unstarred tone are used to capture the slight fall in pitch on this syllable. Next, *graft* is a metrically prominent phrase-final syllable; the pitch is fairly level on this syllable so it warrants a single starred tone. Next, *uh* and *there* are also labeled with metrical prominences. The simple fall to each of the two syllables, which are themselves level in pitch, indicates that a single starred tone is warranted.

Finally, consider how phrase-initial and phrase-final tones are labeled in <<flipside>>. In this example, there is a minor phrasal boundary after *likewise*. The phrase-final syllable -*wise* is not metrically prominent, so it is assigned an unstarred tone with a leftward "+" diacritic, due to the preceding starred tone. The following syllable, *you*, is phrase-initial, but it does not exhibit a local change in pitch; rather, it occurs in the middle of a region of flat pitch which in fact begins on the preceding syllable, -*wise*. As a result, it does not get a tonal marker. Finally, there is a phrase-final syllable at the end of the utterance on -*side*. This syllable is not metrically prominent, so conventions for labeling nonprominent syllables apply. In this case, the tonal movement on the syllable is complex, showing first a fall to a low and then a small rise to a relatively higher pitch. The conventions dictate that this complex tonal sequence should be labeled by two separate unstarred tones.

Labeling practice: Labeling starred, unaccented tones and phrase-final tones.
<<armani4>>
<<armani7>>
<<armani8>>
<<coffee2>>
<<coffee3>>
<<coffee4>>
<<marla>>
<<dugong>>
<<lenient>>
<<institution>>
<<canadians>>
<<baby>>
<<business>>
<<fiscal>>

3.11 Parallelism

Another important phenomenon which affects the labeling of both rhythm and tones is *parallelism*. The term *parallelism* describes a prosodic pattern in which all or part of an intonation contour is repeated, giving rise to parallel sequences of tonal events. Moreover, the phenomenon of parallelism entails that repeated tonal sequences have the

same or similar metrical structure. This section describes parallelism and illustrates it through several examples.

One example illustrating parallelism is <<pushups>>. In this example, the speaker seems to utilize a repeated falling pattern throughout the utterance across one- or two-syllable units. Moreover, there is parallel metrical structure for each falling event. First, consider the fact that each fall can be decomposed into a sequence of a high tone and a low tone. For each fall, the high tone seems metrically strong, while the low tone seems metrically weak. This symmetry in the metrical status of similar tonal events is a general property of tones which are parallel. Note that the the left and right edges of the region over which the parallel repetition occurs is bracketed with "(//" and "//)" markers. These should be placed before the first tone label and after the second tone label which participate in the parallel construction.

Parallelism is also associated with a building up of expectation that a pattern will continue. This perceived repetition can occasionally influence RaP labeling by setting up a repeated pattern and lowering the threshold for hearing an event as repeating the pattern. This is the case in <<pushups>> for the word *like.* This word when considered in isolation does not exhibit much of a fall. However, listeners will tend to hear *like* as having the same metrical and tonal structure as other events in the sequence. As a result, this word is labeled as H* L+, in a manner which parallels other labels in the sequence. RaP encourages the selection of tonal labels which reflect identifiable parallelism with other tonal events in the vicinity. This is justified on the grounds that listeners hear repetition and tend to perceive a pattern as continuing, which is a general phenomenon in auditory perception.

Another example of parallelism can be seen in <<heat>>. In this example, there is a repeated sequence of high and low tones. Moreover, all the high tones are metrically strong, while the low tones are metrically weak. Parallelism is marked on the "misc" tier using "(//" and "//)" labels.

Yet another example of parallelism comes from <<armani3>>. In this example, there is a parallel sequence of a rise to an accented syllable, followed by a fall. Accordingly, the entire utterance is indicated to be parallel, as noted in the "misc" tier by the "(//" and "//)" markers. Note that the utterance-initial tone is labeled :!L+, while the utterance medial tone in a parallel position on *the* is H+. It will often be the case that the initial tone in a parallel sequence does not appear to be repeated in the parallel construction. This is simply a fact about labeling parallelism and arises from the fact that parallel constructions are created through repetition of sequences of *relations*. The first tone in the parallel sequence itself forms a referent for the following tone, so that the initial tone itself in fact constitutes part of the sequence in spite of the identity of its label. Thus, it is important to simply be aware that the first tone in a parallel sequence may often bear a tone label that does not fit with the overall repetition.

Note also that for the purposes of labeling parallelism, all unstarred tones are equivalent. What is important is that similar, corresponding tonal labels have similar metrical structure. In this sense, all unstarred tones are alike in that they all reflect metrical nonprominence. Thus, so long as a tonal sequence shows a repeated pattern, diacritics such as "+"can be treated as equivalent for the purposes of parallelism.

In most situations, parallelism will simply be marked in the "misc" tier. However, in some cases parallelism may influence the labeling in the "tones" or "rhythm" tiers.

One example in which the parallel tone structure influences rhythm labeling comes from <<kansas>>. In this example, the tones and rhythm labeling is fairly straightforward up until the end of the utterance. However, each of the last two words, *last* and *night*, feels like a metrically strong syllable. Recall that syllables in adjacent positions usually should not be labeled as beats. Moreover, *night* clearly seems stronger than *last*, suggesting that one option would simply be to label *night* as a beat and *last* as a nonbeat. How can we decide on the rhythm labeling for the last two syllables?

A key observation is that *last* and *night* have parallel, falling intonation patterns. Recall that parallel intonation involves parallel metrical structure, in the sense that corresponding tones in the sequence have similar metrical status. To label the falling pattern across each syllable requires labeling a sequence of high and low tones, where the high tones are metrically prominent in each case. However, in order for H* tones to be labeled on both *last* and *night*, both syllables must be labeled as beats. In this way, the presence of parallelism can influence the choice of rhythm labels.

Another example of parallelism comes from <<tech_center>>. In this utterance, there is a repeated pattern of stepping down. This is captured through an alternation between metrically prominent tones, all of which are low except for the initial tone, and unstarred equal tones. We mentioned earlier that the initial tone in the parallel sequence need not have a label that fits with the overall pattern, since this tone's contribution to the parallelism is simply to provide a referent for pitch level for the following tone.

Portions of speech which are not adjacent in time can also be parallel; this is referred to here as "long-distance parallelism". An example of this phenomenon comes from <<flipside>>, and the presence of long-distance parallelism again influences the choice of tonal labels. In this example, the initial syllable *like-* has a rising pitch movement consistent with a sequence of a low and a high tone. However, it is unclear initially whether the low tone or the high tone sounds more prominent. Should the transcription be :L+ H* or :L* +H?

The answer is determined by long-distance parallelism between the rising-falling pattern on *likewise* and the one which occurs later in the utterance on *flipside*. Because of the perceived parallelism between *like-* and *flip-*, a transcription on *like-* which has a similar tonal labeling to that of *flip-* is preferred. On *flip-*, the high-pitched event is clearly prominent, so the associated high tone gets a star. This leads to a preference for a transcription on *like-* which also has a high starred tone: :L+ H*. Note that long-distance parallelism is labeled by using the same numeral outside of the parentheses on each instance of the parallel construct.

Another example of long-distance parallelism comes from <<treehouse>>. In this example, the complex rising-falling-rising pattern on *classmate* is repeated on *treehouse*. The corresponding parallel parts are enclosed in "1(//" and "//)1" markers in the "misc" tier. Moreover, *who lives in a treehouse* is echoed downstream in the tonal pattern of *was written up in Atlantic*. These corresponding parts are also marked in the "misc" tier using "2(//" and "//)2" labels. In general, whenever a new pattern is introduced which is echoed downstream, a new number is assigned to the corresponding parallel parts.

This concludes the introductory manual to the RaP labeling system.

Labeling practice: Parallelism.
<<okay>>
<<anne>>
<<my_word>>
<<stretch>>
<<yup>>
<<promiscuity>>

References

Beckman, M., & Ayers-Elam, G. (1997). *Guidelines for ToBI labeling, version 3.0*: The Ohio State University.

Boersma, P. and Weenink, D. (2002) Praat, a system of doing phonetics by computer. Software and manual on line at http://www.praat.org.

Dilley, L. (2005) The phonetics and phonology of tonal systems. Ph.D. thesis, MIT.

Dilley, L., Ladd, D.R., and Schepman, A. (2005) Alignment of L and H in bitonal pitch accents: testing two hypotheses. *Journal of Phonetics*, 33(1), 115-119.

Grice, M. (1995) Leading tones and downstep in English. *Phonology* 12, 183-233.

Halle, M. and Vergnaud, J.-R. (1987) *An Essay on Stress*. Cambridge, MA, MIT Press.

Hayes, B. (1995) *Metrical Stress Theory*. Chicago: University of Chicago Press.

House, D. (1990) Tonal perception in speech. Lund: Lund University Press.

Ladd, D. R. and Schepman, A. (2003) "Sagging transitions" between high accent peaks in English: experimental evidence. *Journal of Phonetics*, 81, 81-112.

Pierrehumbert, J. (1980) *The phonology and phonetics of English intonation.* Ph.D. dissertation, MIT, Cambridge, MA.

Silverman, K. E. A., Beckman, M., Pierrehumbert, J., Ostendorf, M., Wightman, C. W. S., Price, P., & Hirschberg, J. (1992). ToBI: A standard scheme for labeling prosody, *Proceedings of the 2nd International Conference on Spoken Language Processing* (pp. 867-879). Banff.