

# **Embarking on PoLaR Explorations: A Framework for Intonational Annotation and Analysis**

Byron Ahn,  
Nanette Veilleux,  
Stefanie Shattuck-Hufnagel,  
and Alejna Brugos

*draft*  
November 2022



# Contents

|  |            |
|--|------------|
| <b>Contents</b>  | <b>i</b>   |
| <b>List of Figures</b>   | <b>iii</b> |
| <b>List of Tables</b>  | <b>iv</b>  |
| <b>1 PoLaR Guidelines: Advanced Labels</b>   | <b>1</b>   |
| 1.1 PoLaR Advanced Labels . . . . .  | 2          |
| 1.2 PrStr (Advanced) . . . . .   | 3          |
| 1.2.1 Advanced Phrasing Labels . . . . .   | 3          |
| 1.2.1.1 Phrase-Initial Boundaries . . . . .  | 4          |
| 1.2.1.2 Summary of Phrasing Labels . . . . .   | 6          |
| 1.2.2 Advanced Prominence Labels . . . . .   | 6          |
| 1.2.2.1 Summary of Prominence Labels . . . . .   | 7          |
| 1.2.3 Disfluency . . . . .   | 8          |
| 1.2.4 Examples with Advanced PrStr labels . . . . .  | 10         |
| 1.2.5 Summary of the Advanced Labels for the PrStr Tier . . . . .  | 11         |
| 1.3 Points (Advanced) . . . . .  | 12         |
| 1.3.1 Optional Advanced Labels for Relating Objects on Points Tier to Objects on Prosodic Structure Tier . . . . . | 12         |
| 1.3.2 Some Introductory Examples of Advanced Points Labels . . . . .   | 15         |
| 1.3.3 Ambiguous Association with a PrStr Object . . . . .  | 17         |
| 1.3.4 Points associated with phrase-initial boundaries . . . . .   | 18         |
| 1.3.5 More Examples of Advanced Points Tier Labels . . . . .   | 20         |
| 1.3.6 Summary of the Advanced Labels for the Points Tier . . . . .   | 26         |
| 1.4 Ranges (Advanced) . . . . .  | 27         |
| 1.4.1 A complex example of many range changes . . . . .  | 28         |
| 1.4.2 How Many Range Intervals? Different Approaches and Their Implications  | 31         |
| 1.4.3 Ranges and Pitch Floors . . . . .  | 34         |
| 1.4.4 Some Trickier Cases with Local Pitch Ranges and Not-Attested Minimums  | 35         |
| 1.4.5 Sub-ranges and super-ranges: Grouping together ranges using a common floor . . . . .                         | 40         |
| 1.4.6 Summary of the Advanced Labels for the Ranges Tier . . . . .   | 42         |
| 1.4.7 Some English Pitch Range Phenomena . . . . .   | 42         |
| 1.4.7.1 Range Compression: a commonly-occurring local range change   | 42         |
| 1.4.7.2 Expanded ranges: when a range is compressed relative to what follows . . . . .                             | 45         |

|         |   |           |
|---------|---|-----------|
| 1.4.7.3 | A few examples of range changes signalling parentheticals . . . . . | 45        |
| 1.4.8   | How ToBI and PoLaR treat Ranges . . . . .                           | 47        |
| 1.4.8.1 | How ToBI and PoLaR handle downstepping . . . . .                    | 48        |
| 1.4.8.2 | How ToBI and PoLaR handle upstepping . . . . .                      | 49        |
| 1.4.8.3 | Range-labelling challenges addressed by PoLaR . . . . .             | 49        |
| 1.5     | Levels (Advanced) . . . . .   | 50        |
| 1.5.1   | The Influence of Levels and Ranges on One Another . . . . .         | 51        |
| 1.5.2   | Levels and Phonological Categories . . . . .                        | 53        |
| 1.5.3   | Outlook for Levels Annotation . . . . .                             | 54        |
|         | <b>Bibliography</b>   | <b>57</b> |

# List of Figures

|      |  |    |
|------|--|----|
| 1.1  | to_volunteer-halved: An example annotated with PoLaR's Basic labels. . . . .   | 2  |
| 1.2  | The labeller hears the phrase boundary between "obvious" and "educate" as stronger than the phrase boundary between "first" and "seems". . . . . | 4  |
| 1.3  | The labeller is annotating an analysis in which "lion" is not phrase-final, even though "which" is phrase-initial. . . . .                       | 5  |
| 1.4  | The labeller is annotating an analysis in which "selling" is not phrase-final, just because "I" is phrase-initial. . . . .                       | 5  |
| 1.5  | The labeller has annotated the beginning of each phrase as well as its end. . . . .  | 6  |
| 1.6  | The labeller is annotating the perception that "Naomi" and "email" are both prominent, but that "Naomi" is significantly more prominent. . . . . | 7  |
| 1.7  | in_a_plane-disfluencies, with Advanced PoLaR labels. . . . .   | 8  |
| 1.8  | spoon_here, with Advanced PoLaR labels. Note the 'd' on PrStr. . . . .   | 9  |
| 1.9  | experience, with Advanced PoLaR labels. Note the region flanked by '{d' and 'd}' on PrStr. . . . .   | 9  |
| 1.10 | The labeller hears especially strong breaks after "huh" and after "again". . . . .   | 10 |
| 1.11 | The labeller is at several points uncertain about prominence. . . . .  | 10 |
| 1.12 | The labeller hears several disfluencies. . . . .   | 11 |
| 1.13 | amelia_knew2, with Basic PoLaR labels. . . . .   | 13 |
| 1.14 | A grammar of the core Advanced Points tier labels. . . . .   | 14 |
| 1.15 | me, with Advanced PoLaR labels. . . . .  | 16 |
| 1.16 | really_would_not_annoy, with Advanced PoLaR labels. . . . .  | 16 |
| 1.17 | The [i . . . . .   | 18 |
| 1.18 | An initial high-f0 turning point is given a [< label in the Points tier. . . . .   | 19 |
| 1.19 | An initial high-f0 turning point is given the label [</*< label in the Points tier. . . . .  | 20 |
| 1.20 | Advanced Points tier labels here indicate that the labeller associates f0 points with phrase ends or beginnings. . . . .                         | 20 |
| 1.21 | amelia_knew1, with Advanced PoLaR labels. . . . .  | 21 |
| 1.22 | amelia_knew2, with Advanced PoLaR labels. . . . .  | 22 |
| 1.23 | sara_murray, with Advanced PoLaR labels. . . . .   | 22 |
| 1.24 | sara_molina, with Advanced PoLaR labels. . . . .   | 23 |
| 1.25 | menu_eleven, with Advanced PoLaR labels. . . . .   | 24 |
| 1.26 | out_of_order-plateau, with Advanced PoLaR labels. . . . .  | 25 |
| 1.27 | A finite state grammar for Advanced Points tier labels. . . . .  | 27 |
| 1.28 | An idealization of pitch ranges in four hypothetical utterances. . . . .   | 28 |
| 1.29 | An example of four utterances with nine pitch range intervals. . . . .   | 29 |
| 1.30 | A closer look at some Ranges intervals. . . . .  | 29 |
| 1.31 | A single Range interval for the entire utterance. . . . .  | 31 |

|      |   |    |
|------|---|----|
| 1.32 | Each colored band represents the chunk of pitch space corresponding to each numbered Level, 1-5. . . . .  | 32 |
| 1.33 | Each colored band represents the chunk of pitch space corresponding to each numbered Level, 1-5. . . . .  | 33 |
| 1.34 | Parentheses labels are used to mark an advanced intuition that there is a second pitch range within what had been labelled as one pitch range by a labeller using PoLaR Basic labels. . . . . | 34 |
| 1.35 | Pitch maxima seem to vary more than pitch minima, for a given speaker.( <i>From Liberman &amp; Pierrehumbert 1984</i> ) . . . . .   | 35 |
| 1.36 | A hypothetical pitch track. . . . .   | 36 |
| 1.37 | A recording with an attested pitch minimum of about 250Hz to 365Hz. . . . .   | 36 |
| 1.38 | The filled box indicates the speaker's intended local pitch range, and the black dots indicate f0 observations. . . . .   | 37 |
| 1.39 | Labels capturing the intuition that the pitch range minimum (intuited as 135Hz) is not attested in the f0 track. . . . .  | 38 |
| 1.40 | Labels capturing the intuition that the pitch range minimum (intuited as 170Hz) is not attested in the f0 track. . . . .  | 39 |
| 1.41 | The labeller provides an educated guess of the pitch minimum during " <i>out there</i> ". . . . .   | 39 |
| 1.42 | A hypothetical pitch track. . . . .   | 40 |
| 1.43 | A hypothetical pitch track. . . . .   | 40 |
| 1.44 | <i>librivox-zip1</i> , with Advanced PoLaR labels. . . . .  | 41 |
| 1.45 | One prosodic phrase (PrStr annotation) and two local pitch ranges (Ranges annotation). . . . .  | 43 |
| 1.46 | Multiple Ranges intervals, where the labeller's analysis includes multiple instances of range compression. . . . .  | 44 |
| 1.47 | <i>another-little-banana</i> , with Advanced PoLaR labels. . . . .  | 45 |
| 1.48 | An example of both the pitch floor and ceiling of a range shifted upward. . . . .   | 46 |
| 1.49 | An example of a compressed parenthetical. . . . .   | 46 |
| 1.50 | An example of a locally expanded range. . . . .   | 47 |
| 1.51 | An example of a locally compressed range. . . . .   | 47 |
| 1.52 | Different Range labels and their impacts on Levels labels. . . . .  | 51 |
| 1.53 | Interplay of Ranges labels Levels labels. . . . .   | 52 |
| 1.54 | Levels values where the attested f0 minimum differs from the intended pitch minimum. . . . .  | 53 |

## List of Tables

|     |  |    |
|-----|--|----|
| 1.1 | Advanced PrStr phrasing labels: encoding uncertainty or boundary strength. . . . .                 | 6  |
| 1.2 | Advanced PrStr prominence labels: encoding uncertainty or prominence strength. . . . .             | 7  |
| 1.3 | Advanced PrStr labels: encoding disfluency. . . . .  | 10 |
| 1.4 | The PrStr labels described in this chapter, which can be used by Advanced PoLaR labellers. . . . . | 12 |

|     |   |    |
|-----|---|----|
| 1.5 | Advanced Points labels . . . . .  | 15 |
| 1.6 | An expanded set of labels for the Points tier, with several advanced options. . . . . | 26 |
| 1.7 | An expanded set of labels for the Ranges tier. . . . .                                | 42 |

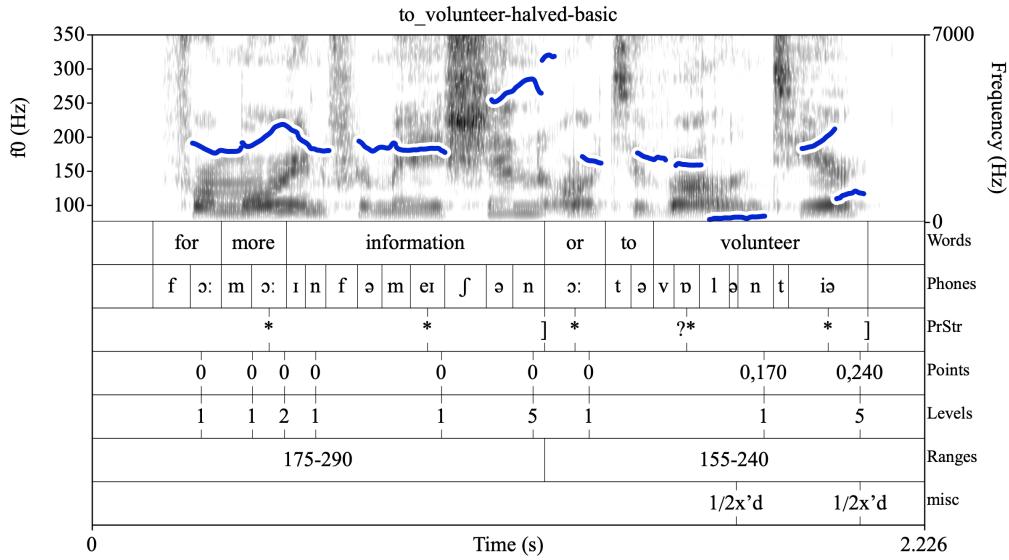


# Chapter 1

## PoLaR Guidelines: Advanced Labels

A distinct advantage of PoLaR is that labellers can become quite proficient in how to do the “Basic” labels with little in the way of training. This is in large part due to the fact that PoLaR *does not* require its labellers to adopt any theory of either how its various labelled characteristics are expected to distribute, or how the characteristics are connected to one another. (This is a departure from other annotation systems, such as those designed to encode contrastive phonological categories.) Thus, a PoLaR labeller need not consider theoretical models or the analytical ramifications of whether to label an aspect of the intonation with a phonological label (or which phonological label to use).

While this design characteristic allows PoLaR labels to be used proficiently (and quickly!) by people who are not working with an eye towards theoretical models, it is still true that many labellers do have theoretical commitments or analyses that they want to encode into their labels. For example, in the example in Figure 1.1, a labeller may wish to annotate that some of the pitch turning points are associated with exponents of a syllable prominence (i.e., prominence-marking) while others are associated with exponents of a phrase boundary (i.e., boundary-marking).



**Figure 1.1:** *to\_volunteer-halved*: An example annotated with PoLaR’s Basic labels.

There is no way to annotate this using PoLaR’s Basic labels, but this is the sort of thing the PoLaR Advanced labelling conventions (described in this chapter) have been designed for. Exactly how this and other analytical moves get transcribed using PoLaR Advanced labels will be described in the sections that follow.

While PoLaR Basic labels are designed to elicit agreement (even among labellers with different backgrounds in prosodic theory), Advanced labels may require labellers to have shared theoretical/analytical priors, in order for agreement to be reached. The PoLaR annotation system itself is not designed to provide answers that can adjudicate in these discussions, but rather it is meant to provide a starting point for these discussions to take place.

Before moving on, let us be clear about the optionality of using PoLaR Advanced labels. As described briefly above, PoLaR Advanced labels serve as a means to systematically encode higher levels of analysis or deeper intuitions that the labeller wishes to commit to. While other annotation systems require such higher-level commitments in most/all of their labels, PoLaR Advanced labels are always opt-in. This means that in cases where the labeller does not wish to commit to a particular analysis, they are not required to do so. (In sum, mixing PoLaR Basic labels with Advanced labels is both expected and encouraged, and researchers are encouraged to adopt the (sub)set of Advanced labels that best fits their research goals.)

## 1.1 PoLaR Advanced Labels

The sections that follow describe what we call PoLaR Advanced labels. These labels occupy the same four core PoLaR tiers (PrStr, Points, Levels, Ranges), but can encode more of the analyses and intuitions of advanced labellers, allowing for the annotation of relationships across tiers and for the inclusion of ideas informed by prosodic theories. In addition, these labels may be especially useful for describing data beyond idealized forms of mainstream varieties of a language produced in a laboratory setting.

Because these labels allow for inclusion of particular analyses, this chapter will have more discussion of how PoLaR labels can connect with prosodic theories and phonological analyses, and also how PoLaR fits in with / compares to existing annotation systems, especially those in the AM tradition. (See also Ahn et al. 2019.)

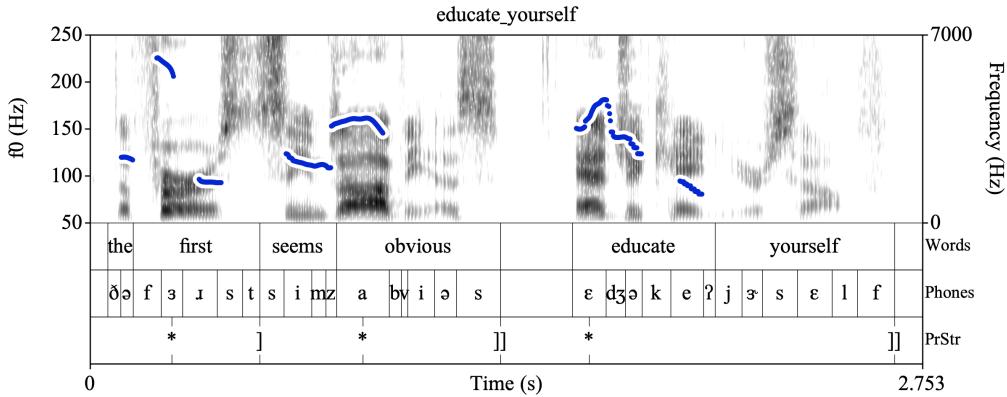
A note before introducing these new PoLaR labels: Even with the addition of Advanced labels, these guidelines will not provide an exhaustive set of labels that could be used in PoLaR. Labellers are encouraged to take advantage of PoLaR's flexibility, by augmenting or changing the labels to fit their needs/views. However, whenever labellers do deviate from the descriptions set out in these guidelines, they need to provide a document that defines the labels they do use, providing descriptions that are at least as clear as those found in these guidelines (preferably with figures showing examples), so that other users have an explicit understanding of how those labels work.

## 1.2 PrStr (Advanced)

In the PrStr tier, phonological objects related to prominence and phrasing are annotated. In the Basic set of labels, the relevant labels are \* and ] (and their “uncertain” partners, ?\* and ?]). One of the ideas behind reducing the number of necessary labels to \* and ] is to simplify the labelling process, allowing even untrained native speakers to provide such labels, in the spirit of Cole’s Rapid Prosody Transcription (Cole et al. 2014). At the same time, it may be desirable (depending on the demands of the research at hand, and the abilities of the labeller(s)) to provide more detailed intuitions, beyond simply indicating the existence of prominence and phrase boundaries. To serve this need, this section will provide an expanded set of labels for the Prosodic Structure tier, for those who wish to include additional information about prominence and phrasing.

### 1.2.1 Advanced Phrasing Labels

A key component of any prosodic analysis is an understanding that there are a number of levels of hierarchical structure. English is commonly thought to make use of two phonologically distinct categories of prosodic phrasing (e.g., Selkirk 1981, Beckman & Pierrehumbert 1986, Nespor & Vogel 1986, *et seqq.*), with evidence for such a distinction being rooted in different sorts of boundary phenomena (having to do with pitch movements, speech rate, pauses, pitch reset, etc.; Brugos et al. 2019). As such, an advanced labeller may wish to distinguish some phrase boundaries from other phrase boundaries that are stronger. Thus in addition to the default ] label, PoLaR provides the ]] label for phrase boundaries that are perceived as stronger than others.

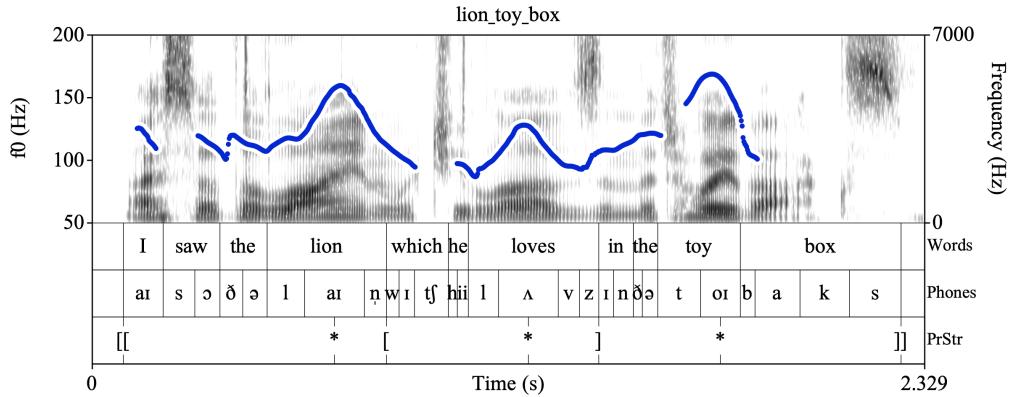


**Figure 1.2:** The labeller hears the phrase boundary between “obvious” and “educate” as stronger than the phrase boundary between “first” and “seems”.

PoLaR does not require the labeller to adhere to any particular rigorous definitions for different sorts of prosodic phrases, across recordings. It may simply be that a labeller wishes to distinguish different levels of phrase-sizes in a particular example (perhaps due to their intuitive perception of disjunction, or perhaps because some acoustic cue such as length of pause is particularly strong). Though labellers need not adhere to any particular rules on when to use ] versus ]], labellers may find it helpful to annotate in a “misc” tier what makes a ]] phrase boundary stand out as especially perceptually strong.

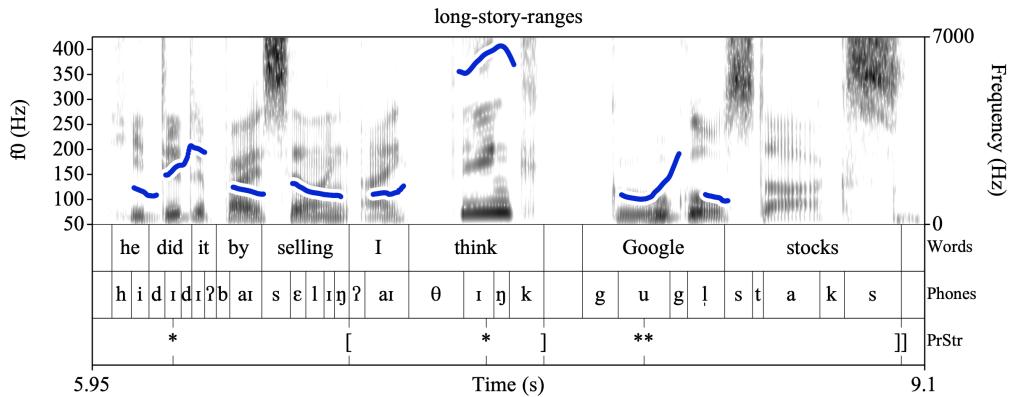
### 1.2.1.1 Phrase-Initial Boundaries

In addition to different sorts of phrase-final boundaries, a labeller may wish to also label the left edges of phrases, by using parallel labels like [ , ?[ , and [ [. This may be useful if operating under an approach in which not every right edge of an intonation phrase entails the left edge of another intonation phrase. Because the Strict Layer Hypothesis (cf. Nespor & Vogel 1986) in effect *prohibits* recursion in prosodic structure, annotating a right edge is essentially annotating a left edge. However, many strands of contemporary research *allow* recursion in prosodic structure (e.g., Ladd 2008, Selkirk 2011, Itô & Mester 2012, and Kentner & Féry 2013), making it useful to annotate both left and right edges of a phrase. For example, in a case of center-embedding like in “*I saw the lion which he loves in the toy box*” with the non-restrictive relative clause, some may analyze the phrasing as “(*I saw the lion (which he loves) in the toy box*)” – i.e., there is a phrase-initial edge before *which* but no phrase-final edge after *lion*. This could be annotated using Advanced phrasing labels as in Figure 1.3.



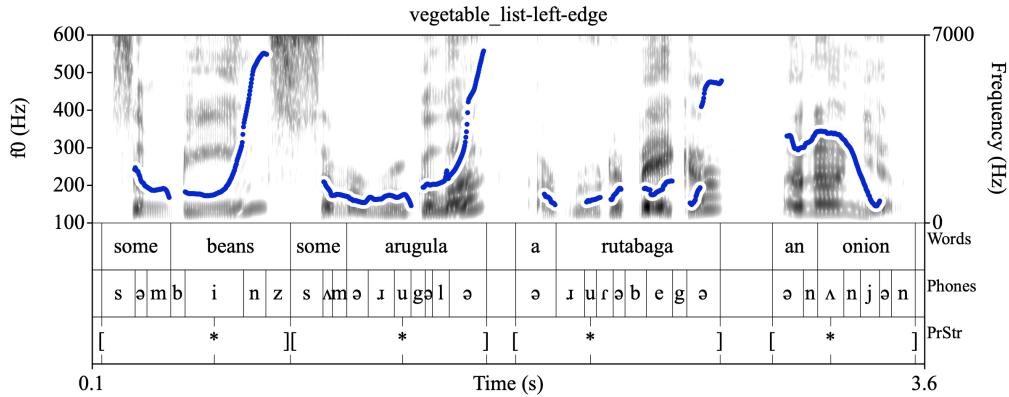
**Figure 1.3:** The labeller is annotating an analysis in which “*lion*” is not phrase-final, even though “*which*” is phrase-initial.

Similarly, with parentheticals that are produced like “*I think*” in Figure 1.4, there are cues to “*I*” being phrase-initial (glottalization), but not many cues to “*selling*” being phrase final. This could be analyzed in terms of nested prosodic phrases, and such an analysis is labelled on the PrStr tier in that figure.



**Figure 1.4:** The labeller is annotating an analysis in which “*selling*” is not phrase-final, just because “*I*” is phrase-initial.

For researchers who choose to label phrase beginnings with the [ label, there may be occasions where a single word boundary coincides with both the end of one phrase and the beginning of another, with no intervening silent interval. It is possible to use ] [ at such points, as is illustrated in Figure 1.5 below:



**Figure 1.5:** The labeller has annotated the beginning of each phrase as well as its end, including an instance of the same location being both a phrase end and a phrase beginning, i.e. the marker between “beans” and “some.”

Note that we offer this [ label as a tool for those who would find using it to be helpful for their research purposes, and using it is by no means required. In other words, use the [ label if you are compelled to, but we take no stance on how often labellers should be using it (if at all). For some cases where [ labels may be useful for Advanced Points labels, see section 1.3.4.

### 1.2.1.2 Summary of Phrasing Labels

Basic labels are highlighted, to be used by all, with the remaining Advanced labels provided for labellers who have more experience or who are targeting particular phenomena. As a reminder: phrasing is cued by acoustics, but these labels are about the labeller’s *perception* or phrasing, in the given utterance context.

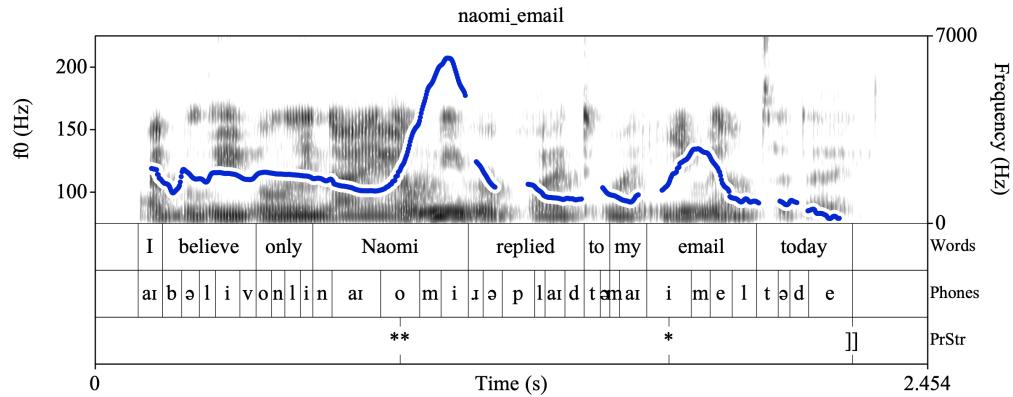
| Label | Phonological Object               | Significance of Label  |
|-------|-----------------------------------|--|
| ]     | Prosodic Phrase Boundary          | A prosodic phrase ends at this point                                     |
| ?]    | Possible Prosodic Phrase Boundary | A prosodic phrase may end at this point, but the labeller is uncertain   |
| ]]    | Large Prosodic Phrase Boundary    | A noticeably large prosodic phrase ends at this point                    |
| [     | Prosodic Phrase Boundary          | A prosodic phrase begins at this point                                   |
| ?[    | Possible Prosodic Phrase Boundary | A prosodic phrase may begin at this point, but the labeller is uncertain |
| [[    | Large Prosodic Phrase Boundary    | A noticeably large prosodic phrase begins at this point                  |

**Table 1.1:** Advanced PrStr phrasing labels: encoding uncertainty or boundary strength.

### 1.2.2 Advanced Prominence Labels

Just as there are levels of prosodic phrasing, in some cases different levels of prosodic prominence can be identified perceptually and/or acoustically. In such cases an advanced labeller may wish to distinguish some prominences from other prominences that are stronger. Thus in

addition to the default \* label, PoLaR provides the \*\* label for prominences that are perceived as stronger than others.



**Figure 1.6:** The labeller is annotating the perception that “*Naomi*” and “*email*” are both prominent, but that “*Naomi*” is significantly more prominent.

As with all Advanced labels, there is no need to use \*\* in every utterance. As default guidance for distinguishing \* and \*\*, labellers should only use the \*\* label when it is very clear that one of the prominences is perceived as especially strong, as compared to the other words in the context of the utterance.

Additionally, as with the ] and ]] labels, PoLaR does not require the labeller to adhere to any particular rigorous definitions for different sorts of prominence across their entire corpus; a labeller may simply wish to distinguish different degrees of prominence in a given example (perhaps as a simple matter of perceptual strength, perhaps because of especially high f0 peaks, etc.). That said, labellers are encouraged to annotate in a “misc” tier what makes a \*\* prominence stand out as especially perceptually strong.

### 1.2.2.1 Summary of Prominence Labels

Basic labels are highlighted, to be used by all, with the remaining labels proposed for labellers who have more experience or who are targeting particular phenomena. As a reminder: prominence is cued by acoustics, but these labels are about the labeller’s *perception* or prominence, in the given utterance context.

| Label | Phonological Object          | Significance of Label  |
|-------|------------------------------|--|
| *     | Prominence                   | The constituent this occurs in is prominent                                      |
| ?*    | Possible Prominence          | The constituent this occurs in might be prominent, but the labeller is uncertain |
| **    | Especially Strong Prominence | The constituent this occurs in is especially prominent                           |

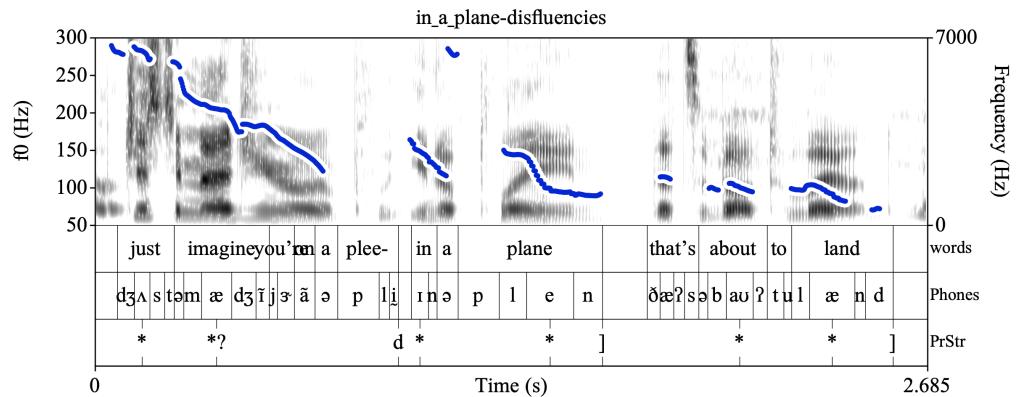
**Table 1.2:** Advanced PrStr prominence labels: encoding uncertainty or prominence strength.

### 1.2.3 Disfluency

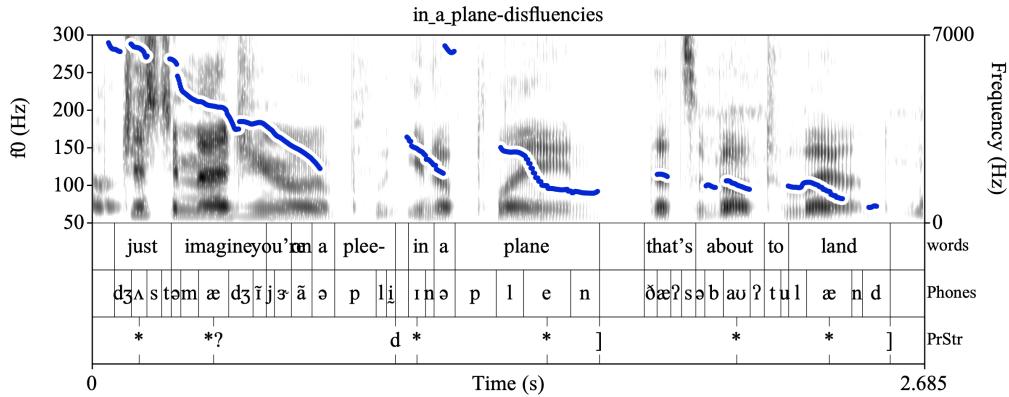
Here we pivot from labelling phrasing and prominence to labelling a phenomenon that may interact with perceived phrasing/prominence: disfluencies in speech. A disfluency is defined here as a disruption of fluent, connected speech, which manifests acoustically as a change in speech rate, a change in pitch trajectory, addition of a pause, etc. Such an acoustic disruption can be momentary in nature, or it can be observed over a period of time. (See Shattuck-Hufnagel & Cutler 1999 for a discussion of how different types of disfluency can be marked prosodically. In addition, Brugos et al. 2019 describes in more detail the range of disfluency effects on prosody that are attested in English.) Characteristic of all types of disfluencies is that, if a speaker were to repeat an idealized form of the disfluency-containing utterance, with the same intended meaning, the idealized repetition could lack all disfluent regions.

As disfluencies can influence prosody in measurable ways, a labeller may find it useful to label disfluent speech as such. For this purpose, PoLaR provides the ‘d’ label, which is to be marked on the PrStr tier, at the point in the utterance where there is a disfluency: i.e., in the word or partial word (in the case of a cut-off word) where the disfluency is noticed.

While disfluencies may differ prosodically from one another, this ‘d’ label can be used for all types. An example of a disfluency that corresponds to a momentary stutter (with no sense of disjunction) is given in Figure 1.7. Additionally, an example of a disfluency that induces a sense of disjunction like that of a phrase boundary (due to the overlap in set of cues) is given in Figure 1.8. In both cases, the ‘d’ label is on the PrStr tier, time-aligned with the onset of the disfluent moment/region, or where the speaker cuts themselves off.

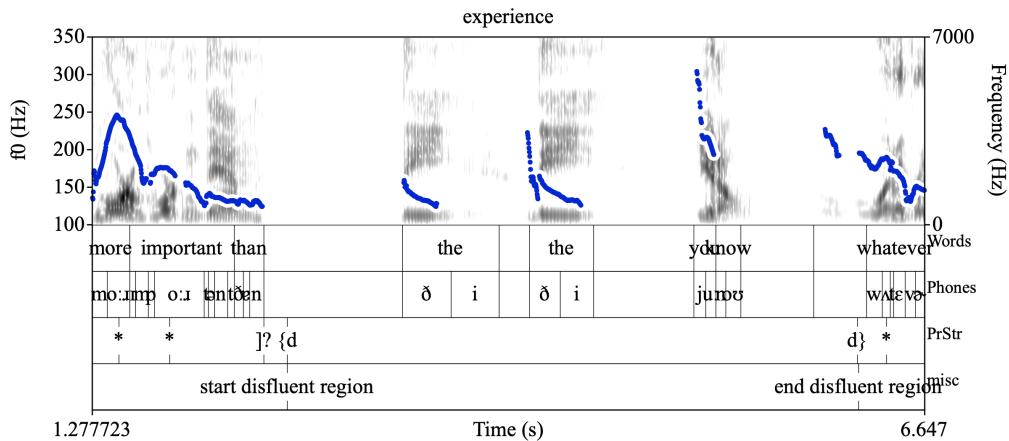


**Figure 1.7:** *in\_a\_plane-disfluencies*, with Advanced PoLaR labels. Note the ‘d’ on PrStr, annotating at the cut-off, before a restart begins.



**Figure 1.8:** spoon\_here, with Advanced PoLaR labels. Note the ‘d’ on PrStr, annotating where a labeller may perceive some of the prosody associated with this disfluency as cues to phrase break.

There may be cases in which a speaker produces not just a single disfluency, but a disfluent region. For such cases, the labeller can designate the beginning and ending of such a region by using the ‘{d’ and ‘d}’ labels in the PrStr tier at the beginning and end (respectively) of this region. The labeller can choose to skip over other labelling decisions within this designated region. Such an example is shown in Figure 1.9.



**Figure 1.9:** experience, with Advanced PoLaR labels. Note the region flanked by ‘{d’ and ‘d}’ on PrStr, which annotate a stretch of time that sounds disfluent.

In all cases, the ‘d’ label of PoLaR allows the labeller to annotate the presence of one or more disfluencies, without any information about the acoustic details or which (if any) type of speech error it corresponds to. Labellers who wish to label more about the details of the disfluency are encouraged to use a system of labels designed to annotate this information, perhaps on a different tier. (See Brugos et al. 2019 for some ideas of disfluency labels that could work alongside PoLaR.)

| Label | Phonological Object         | Time-Aligned with  |
|-------|-----------------------------|--|
| d     | Disfluency                  | A point in the utterance where there is a disfluency               |
| {d}   | Start of a disfluent region | A point in the utterance where a disfluent region of speech begins |
| d}    | End of a disfluent region   | A point in the utterance where a disfluent region of speech ends   |

Table 1.3: Advanced PrStr labels: encoding disfluency.

#### 1.2.4 Examples with Advanced PrStr labels

In this section we have reviewed some ways to systematically label these more refined intuitions, beyond the simple contrast of presence/absence of \* and ] labels; e.g., using the contrast between \* and \*\* labels, and ] and ]] labels to differentiate degrees of prominence/phrasing. In addition to the examples provided earlier in this section, these more fine-grained contrasts are exemplified in Figure 1.10 and Figure 1.11.

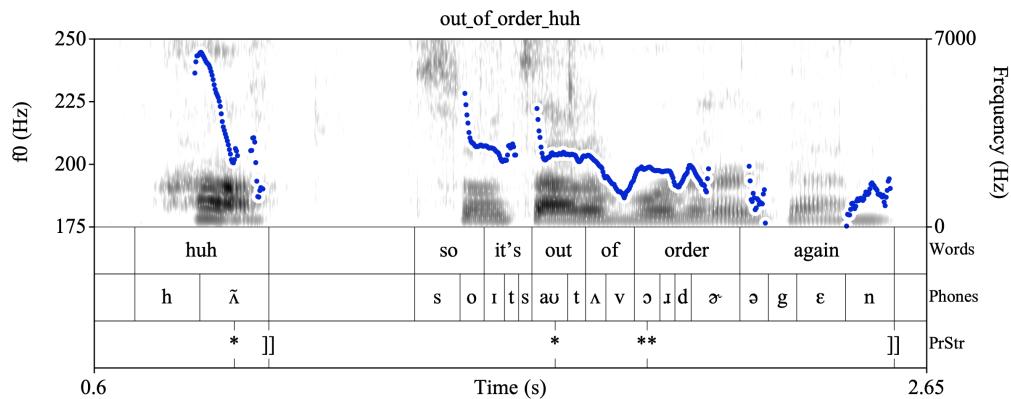


Figure 1.10: The labeller hears especially strong breaks after “huh” and after “again”; they also hear especially strong prominence on “order”.

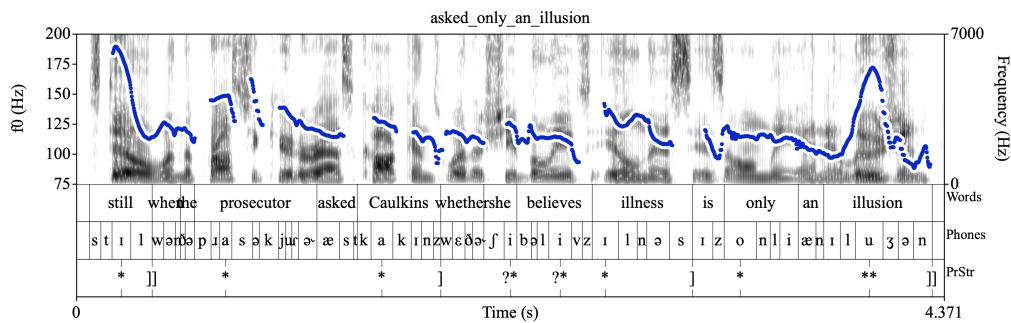
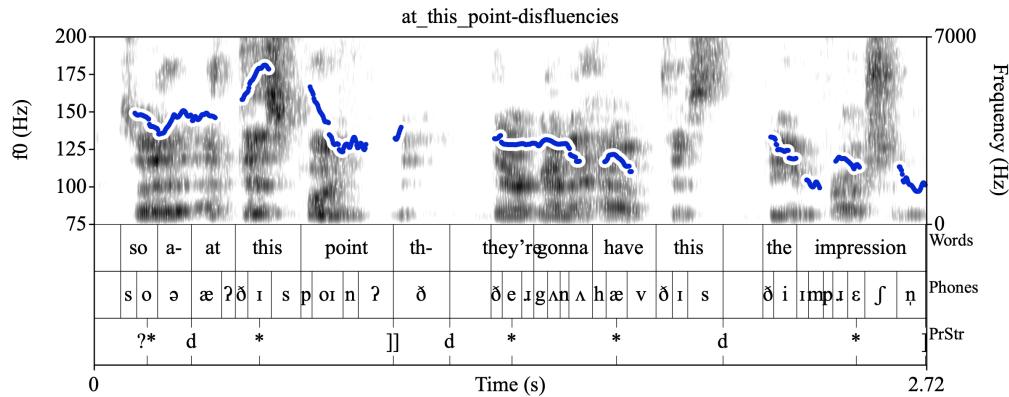


Figure 1.11: The labeller is at several points uncertain about prominence (“she”, “believes”, “only”), while also hearing especially strong prominence on “illusion” and especially strong phrase boundaries after “still” and “illusion”.

The following example Figure 1.12 includes instances of both labeller uncertainty about prominence ( $?^*$ ) and a perceived a disfluency (marked with the 'd' label). Note that some disfluencies have pauses immediately following and some do not; if it feels like the disfluency leads to new prosodic phrase boundaries, these boundaries should be annotated separately. (Likewise for disfluency-induced prominence.)



**Figure 1.12:** The labeller hears several disfluencies, where the speaker restarts/recasts their speech; the 'd' labels that indicate this are at the cut-off point, before any existing disfluency-related pauses.

### 1.2.5 Summary of the Advanced Labels for the PrStr Tier

This section has reviewed the Advanced labels set out for transcription in the PrStr tier. As with all Advanced labels described in this chapter, their significance and guidelines for usage are provided here, but none of the Advanced labels are required for PoLaR labellers. For example, as mentioned for the phrase-initial boundary label, [, some labellers might want to use this label, while others may use it infrequently, or not at all. Similar can be said for the \*\* and disfluency-related labels.

The exhaustive list of Advanced PrStr labels are laid out in the table below, which compiles the tables found throughout this section.

| Label | Phonological Object               | Time-Aligned with   |
|-------|-----------------------------------|---|
| ]     | Prosodic Phrase Boundary          | The end of a Words interval where a prosodic phrase ends  |
| ?]    | Possible Prosodic Phrase Boundary | The end of a Words interval where a prosodic phrase may end (but the labeller is uncertain)         |
| ]]    | Large Prosodic Phrase Boundary    | The end of a Words interval where a noticeably large prosodic phrase ends                           |
| [     | Prosodic Phrase Boundary          | The beginning of a Words interval where a prosodic phrase begins                                    |
| ?[    | Possible Prosodic Phrase Boundary | The beginning of a Words interval where a prosodic phrase may begin (but the labeller is uncertain) |

| Label | Phonological Object          | Time-Aligned with   |
|-------|------------------------------|---|
| [ [   | Large Prosodic Boundary      | The beginning of a Words interval where a noticeably large prosodic phrase                      |
| *     | Prominence                   | The center of the Phones/Words interval that is prominent                                       |
| ?*    | Possible Prominence          | The center of the Phones/Words interval that might be prominent (but the labeller is uncertain) |
| **    | Especially Strong Prominence | The center of the Phones/Words interval that is especially prominent                            |
| d     | Disfluency                   | A point in the utterance where there is a disfluency  |
| {d    | Start of a disfluent region  | A point in the utterance where a disfluent region of speech begins                              |
| d}    | End of a disfluent region    | A point in the utterance where a disfluent region of speech ends                                |

**Table 1.4:** The PrStr labels described in this chapter, which can be used by Advanced PoLaR labellers.

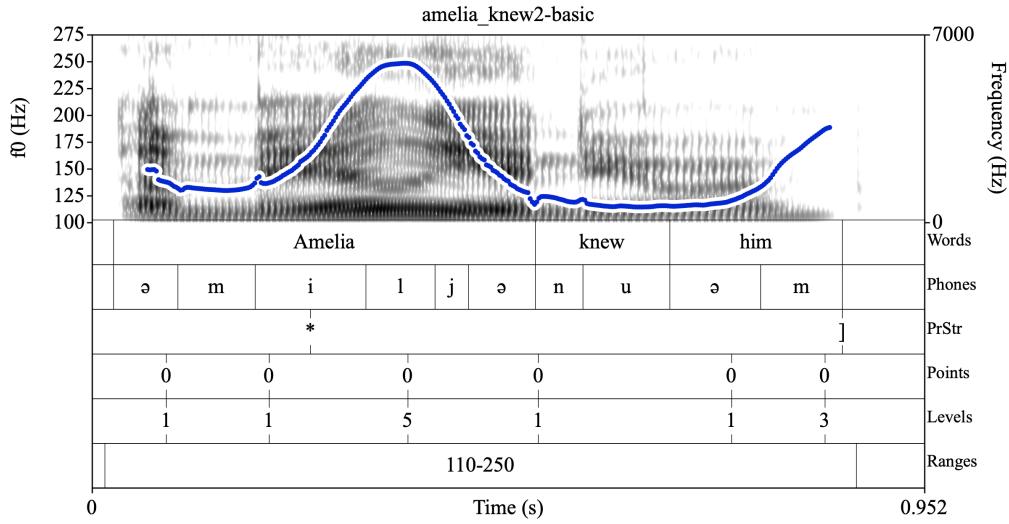
## 1.3 Points (Advanced)

### 1.3.1 Optional Advanced Labels for Relating Objects on Points Tier to Objects on Prosodic Structure Tier

One valuable aspect of the PoLaR system is that it allows labellers to annotate aspects of the intonational phonetics (e.g., time alignment of turning points) separately from aspects of the intonational phonology (e.g., prominences or boundaries). This independence can be achieved through using ‘0’ labels on the Points tier, as we described for Basic Points labels in Chapter [?REF?](#). These PoLaR Basic labels (consisting of ‘0’ labels and optional comma-overrides) can be revised in a second-pass of labelling: in the first pass, PoLaR Basic Prosodic Structure and Points tiers are labelled, and in the second pass, ‘0’ labels are replaced with more complex labels (described in this section) to capture relationships between the labels on the Points tier with those on the PrStr tier. To understand these complex labels, we describe first what such labels are designed to capture, before going into the details of how to use them.

It is commonly understood that some pitch movements (like those annotated on the Points tier) are associated with (i.e., manifest cues to) particular prosodic events (like those annotated on PrStr). For example, some pitch movements are used to mark prominences, while others are used to mark boundaries. The labeller, therefore, might ask themselves if a particular turning point contributes to an intonational contours’s shape that in turn contributes to a listener’s perception of prominence or phrasing. In cases where the answer is ‘yes’, the Advanced Points labels provide the means for the labeller to capture this relationship.

Even relatively naive intonational annotators will note that pitch patterns often serve as acoustic cues to a prominence or phrase boundary. Consider the example in Figure 1.13, annotated with Basic PoLaR labels.



**Figure 1.13:** amelia\_knew2, with Basic PoLaR labels.

Here the third label on the Points tier (marking the f0 peak) is cueing the prominence on “Amelia” labelled on the PrStr tier, while the final two labels on the Points tier (marking the phrase-final rise) cue the utterance-final phrase boundary labelled on the PrStr tier. Which, if any, of the other Points labels are related to the PrStr objects is subject to some debate, and labellers may well disagree.

Users of other annotation systems are intimately familiar with this idea; much of what many other annotation systems do is model the relationship between pitch movements and other prosodic objects. Systems that do this include those that make an important distinction between pitch movements of various types; e.g., IPO, ToBI, IViE, or RaP. Because of this, such labellers have strong intuitions whether a particular f0 turning point “belongs to” a prominence (i.e., manifests a cue for a pitch accent) or a phrase boundary (i.e., manifests a cue for an edge tone). PoLaR labellers who are familiar with those models are encouraged to use their background in these other models to inform how they relate Points and PrStr objects with Advanced Points labels. (Note this means users with different theoretical priors may well label the same Basic Points labels of a single utterance with different Advanced Points labels - we take this to be a feature, not a bug.)

Below we provide a more comprehensive description of the Advanced PoLaR labels that can replace pitch turning points that are originally marked with a ‘0’ on the Points tier, to indicate that a given turning point is perceived to be prominence-marking or edge-marking (or both). A PoLaR Advanced Points label consists of a minimum of two components: (i) the *type* of phonological object the turning point is associated with, and (ii) a *pointer* to the relevant Prosodic Structure tier label’s time coordinate. We find that there are three different scenarios that arise when deciding whether a turning point is associated with a prosodic event labelled on the PrStr tier, and if so, with which one:

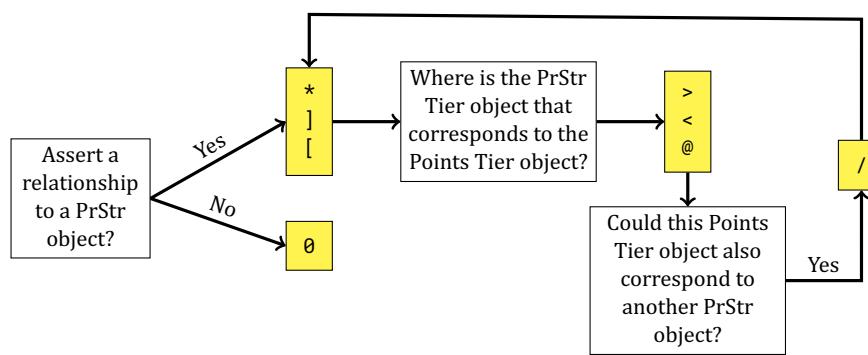
1. No PrStr labelled event is associated with the f0 turning point, or it’s not clear that one is. In this case, use ‘0’ as the complete label. (This is a ‘neutral’ label, indicating that the labeller is not committing to any particular analysis, but has determined that a

- significant turning point needs to be marked at that location.)
2. A particular *type* of PrStr labelled event (either '\*' or ']'; marked below as '\_') is associated with the f0 turning point, which is time-aligned in various ways, indicated by one of three *pointers*:
    - a. PrStr label follows Points label ('\_>')
    - b. PrStr label precedes Points label ('\_<')
    - c. [uncommon] PrStr label is precisely co-timed with Points label ('\_@')
  3. Occasionally, multiple PrStr labelled events are associated with a single f0 turning point. That is, the turning point is either (i) ambiguous between being associated with the preceding PrStr labelled event and being associated with the following event, or (ii) the labeller believes that the point relates to both of those events. In these cases a complex label is used, with two labels separated by a / ('\_</\_>').

There are two observations to be made about these labels right away. First, any non-'0' label on the Points tier indicates the labeller has an intuition about the appropriate analysis. Thus, in the absence of such an intuition, the advanced labeller should leave the '0' label in place.

Second, the pointers themselves ('>', '<', '@') carry no significance beyond indicating a relationship of how a PrStr label and a Points label happen to be time-aligned. That is, because the timing of a PrStr label is at the discretion of the labeller, its precise location in time is not meaningful. Thus, a label '\*>' in the Points tier simply means that "this turning point manifests a cue to the prominence labelled as a '\*' on the Prosodic Structure tier, which is time-aligned later than (to the right of) this Points tier label." A minimally different label, '\*<', says that the relevant PrStr '\*' label is time-aligned earlier than the Points tier label. To repeat, in the interest of clarity: these two labels are not in and of themselves informatively different, in part because the particular timing of the PrStr label is not informative.<sup>1</sup>

The options and the order of the notation is summarized below. (A more complete grammar for Points tier labels is provided in section 1.3.6.)



**Figure 1.14:** A grammar of the core Advanced Points tier labels.

In summary, these Advanced Points tier labels are either (1) the simple '0' label, (2) a complex

<sup>1</sup>In other words, the difference between, e.g., '\*>' and '\*<' is not a meaningful one; it may simply be a reflection of how the '\*' label happens to be timed – which is at the whim of the labeller. (Recall that the '\*' label on the PrStr tier need only be time-aligned within the stressed vowel.) This may be somewhat surprising for labellers familiar with other annotation systems, where labels that use different symbols in a label are meant to signify meaningful/phonological distinctions (e.g., L+H\* vs. L\*+H in ToBI or H\*L vs. !H\*L in IVIE).

label with two components in the order `<type> <pointer>`, or (3) a label made up of two complex labels separated by a `/`.

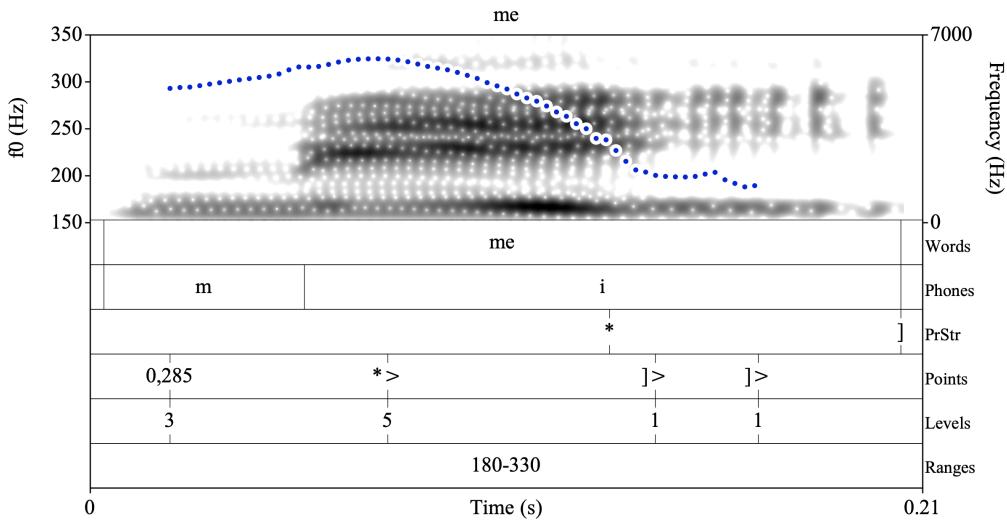
| Type Label    | Relationship with a Prosodic Structure Tier Object  |
|---------------|---|
| 0             | The default label is still the most appropriate choice if the annotator chooses not to affiliate a turning point with a Prosodic Structure label. (No pointer is used in conjunction with a '0' label.) |
| *             | An f0 turning point for a * (of any type; e.g., also ** or ?*) in the Prosodic Structure tier   |
| ]             | An f0 turning point for a ] (also ?] or ]) in the Prosodic Structure tier   |
| [             | An f0 turning point for a [ (also ?[ or [ [ in the Prosodic Structure tier  |
| Pointer Label | Timing of the relevant Prosodic Structure Tier Object   |
| >             | This f0 turning point is related to a Prosodic Structure tier object that follows it  |
| <             | This f0 turning point is related to a Prosodic Structure tier object that precedes it   |
| @             | This f0 turning point is related to a Prosodic Structure tier object that is precisely aligned with it  |
| Special Label | Noted ambiguity   |
| /             | This f0 turning point's affiliation is ambiguous; it may be related to the Prosodic Structure event before the turning point, or to the Prosodic Structure event that follows it.                       |

**Table 1.5:** Advanced Points labels: The possible types of phonological object that a turning point is related to, and the pointers to the location of the Prosodic Structure Tier object in time. By default, a labeller does not offer a phonological analysis, using the label that is highlighted.

### 1.3.2 Some Introductory Examples of Advanced Points Labels

A simple Advanced Points label is shown in Figure 1.15. This is a more complete label of this file, which is shown earlier with its Basic Point labels as Figure [?REF?](#) in Chapter [?REF?](#)). The '`*`' on the Points tier is pointing rightward to the '`*`' label on the PrStr tier that follows it, indicating the labeller's judgment that this turning point helps to cue the prominence marked by the '`*`'.<sup>2</sup>

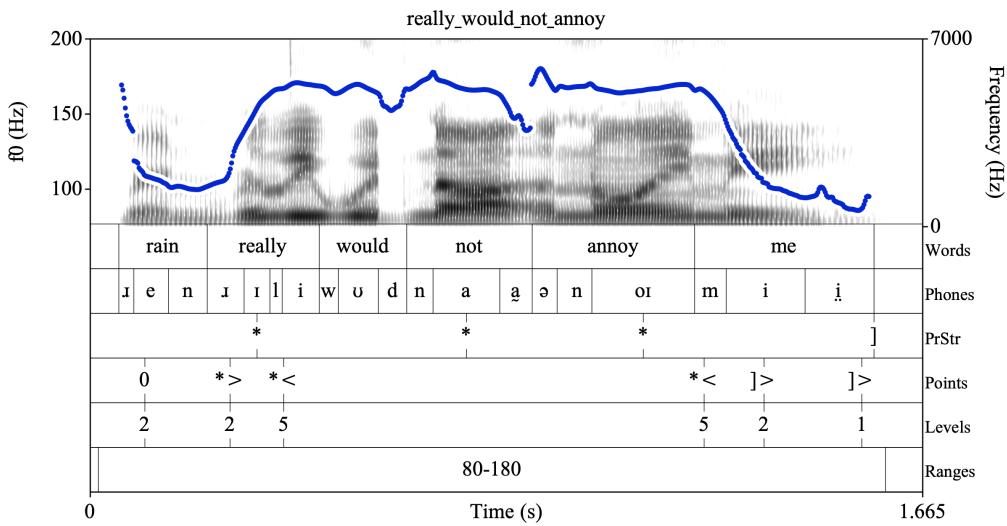
<sup>2</sup>Here the '`*`' label on the Prosodic Structure tier is time-aligned to the precise middle of the stressed vowel interval (as determined by a forced-aligner). This illustrates the general principle that no Prosodic Structure tier object is aligned in any particular way with respect to f0 movements.



**Figure 1.15:** *me*, with Advanced PoLaR labels.

This example also shows the usage of the '] $>$ ' label. This indicates that the f0 movement labelled by the third and fourth points tier labels (formerly '0's) are related to the following phrase boundary, labelled with a  $]$  on the Prosodic Structure tier. Note that there are two of these labels for this recording: one to mark the beginning of the low flat f0 that signals the end of the phrase (cf. ToBI phrase accents), and one to mark the end of that plateau. Also note that the final '] $>$ ' label is not timed to be at (or even near to) the PrStr ' $]$ ', but rather to label the end of the reliable f0 track.

A second example, shown below, was also presented as Figure **?REF?** during the discussion of Basic '0' Point tier labels in Chapter **?REF?**.



**Figure 1.16:** *really\_would\_not\_annoy*, with Advanced PoLaR labels.

Again, note that the Points tier labels ignore unreliable pitch tracks (such as the very high pitch

tracking at the beginning of “*rain*”, the dip in the [d] of “*would*”, the perturbations between “*not*” and “*annoy*”, etc.). For further discussion of such phenomena, the reader is referred to Chapter ?REF? (specifically section ?REF?), for its discussion of pitch track disruptions (a.k.a. “microprosody”).

The first Point tier label, therefore, occurs where the initial f0 becomes reliable. Because that turning point is not perceived to contribute to any of the Prosodic Structure (PrStr) events, the ‘0’ label remains. The second and third Point tier labels are determined to be related to the prominence on the word “*really*” and so the default ‘0’ label is replaced with a \* to indicate its relationship to a prominence (rather than a phrase break) and the ‘>’ and ‘<’ pointers further indicate which prominence: the one labelled in the middle of the vowel in the first syllable of “*really*” in the Prosodic Structure Tier above.

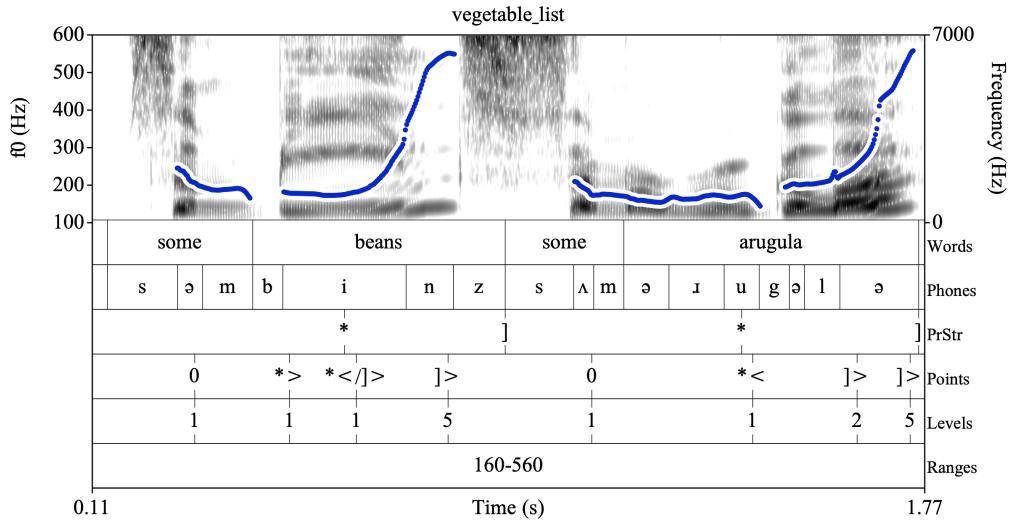
Notably, “*not*”, though prominent, is not marked by any label on the Points tier; it occurs in the middle of a high plateau. To emphasize: not every prominence or phrase boundary in PrStr necessarily corresponds to a Points tier label. The fourth Points tier label is determined to be related to the prominence on “*annoy*”, marked at an earlier time point in the PrStr tier. Finally, the last two Points tier labels are associated with the final low phrase break to their right.

A broader point illustrated by this example is that there is no minimum or maximum number of Points tier objects that are related to a given PrStr tier object. Each of the three \* labels in this example is related to a different number of Points labels (in temporal order, the PrStr \*s are affiliated with two, zero, and one Points labels, respectively).

### 1.3.3 Ambiguous Association with a PrStr Object

There are situations where a turning point in the f0 contour might reasonably be analyzed as associated with multiple prosodic elements. This could be a situation of indeterminacy: i.e., the labeller does not have enough information to decide one way or another. Alternatively, a labeller might have reason to believe that there are two turning points, at an abstract level of the speaker’s intentions, but the acoustic signal is such that only one turning point is identifiable.

As an example, if a prominent syllable is phrase-final, a turning point could be associated with the \* or the ] on the PrStr tier, both of which are labelled on the same syllable. This is illustrated by the word *beans* in Figure 1.17 below.



**Figure 1.17:** v

owel in *beans* is prominent and is also phrase-final.] The [i] vowel in *beans* is prominent and is also phrase-final. The low f0 could be associated with the \* (i.e., a low pitch accent), or it could also be associated with the ] (i.e., as the beginning of a final rise edge tone).

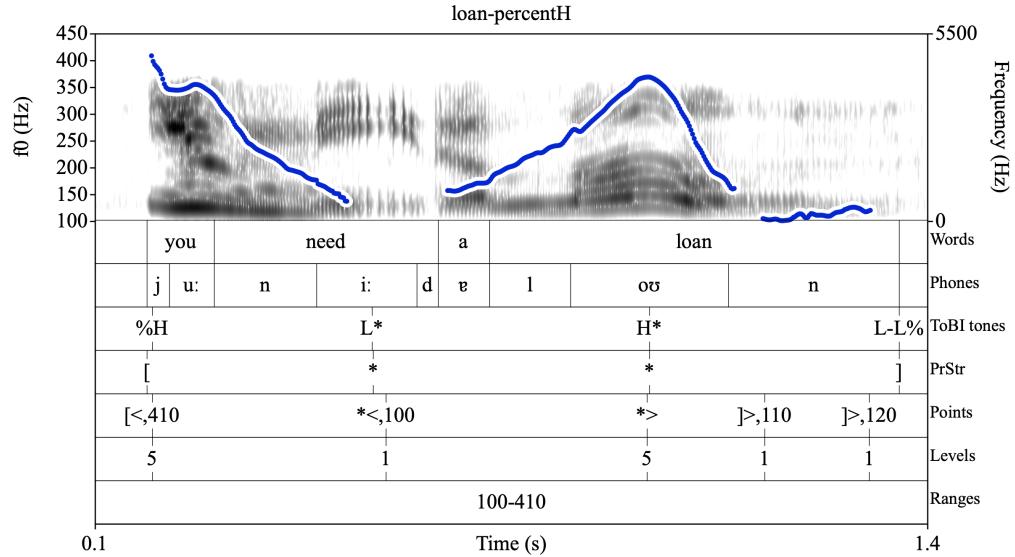
Rather than require the labeller to arbitrarily assign the turning point label leftward or rightward, the / can be used between two labels, allowing the labeller to indicate that they believe both PrStr associations are plausible. This is the case in Figure 1.17, where ‘\*</ ]>’ means the turning point could be associated with the ‘\*’ on the PrStr tier that precedes the Point or with the ‘]’ on the PrStr tier that follows it. Listeners hear “the same” intonational contour being relevant for both *beans* and *arugula*; note the similarities between the two, in particular with respect to their Points labels. While there are two Points labels annotated in *beans* and three in *arugula*, the analysis in them is the same: there is a low point related to the prominence ‘(\*<)’, another low(ish) points related to the boundary (‘]>’), and a high point related to the boundary (‘[ ]>’). It just happens that in *beans*, two of these relationships are annotated at the same point in time (‘\*</ ]>’). The additional syllables in *arugula* (particularly between the prominent [iu] and the phrase-final boundary) provide enough time to see more clearly the motivation for each of the prominence and boundary being associated with a “low” point in the intonation.

Similar to this usage of ‘\*</ ]>’, the / can be used to create compound labels like ‘\*</ \*/>’, ‘\*/> / ]>’, ‘] </ \*/>’, etc. Additional examples with compound labels like this will be found in the following sections.

### 1.3.4 Points associated with phrase-initial boundaries

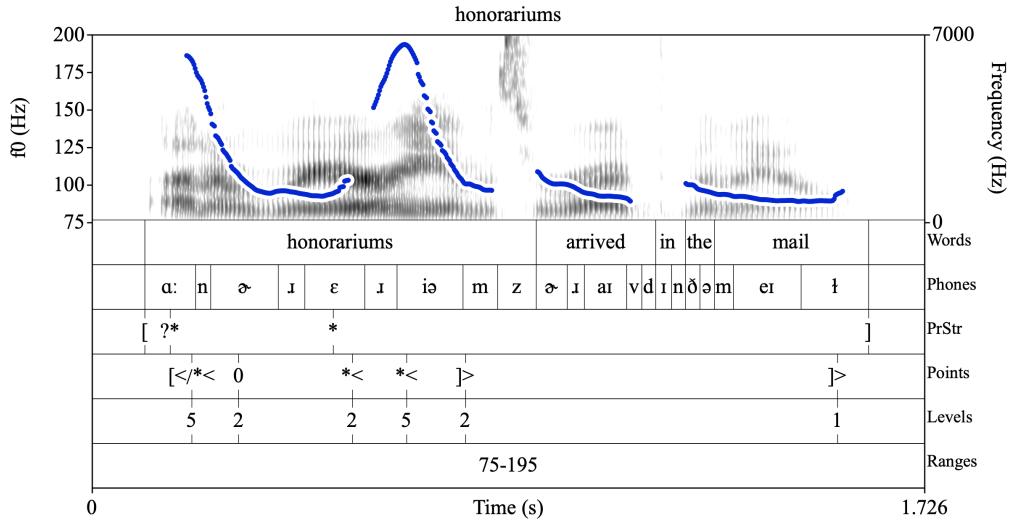
While AM annotation conventions typically annotate only the boundary at the end of a phrase, PoLaR allows the flexibility of optionally marking the beginning of a phrase. In such cases, there may be f0 turning points that the labeller would like to attribute to an initial boundary. A widely-known circumstance in which this may occur is for a high initial f0 in a phrase that is not clearly associated with a prominence. (I.e. those rare cases that are labelled %H in ToBI.) With PoLaR, the labeller may choose to assign any or all phrase-initial turning points to a [ boundary marker.

The example below shows a case where an initial high f0 has been attributed to an initial High boundary tone (%H) in ToBI training materials. In PoLaR, such a case could be labelled with a [ in the PrStr tier, and an early high f0 point in the Points tier with a [< label:



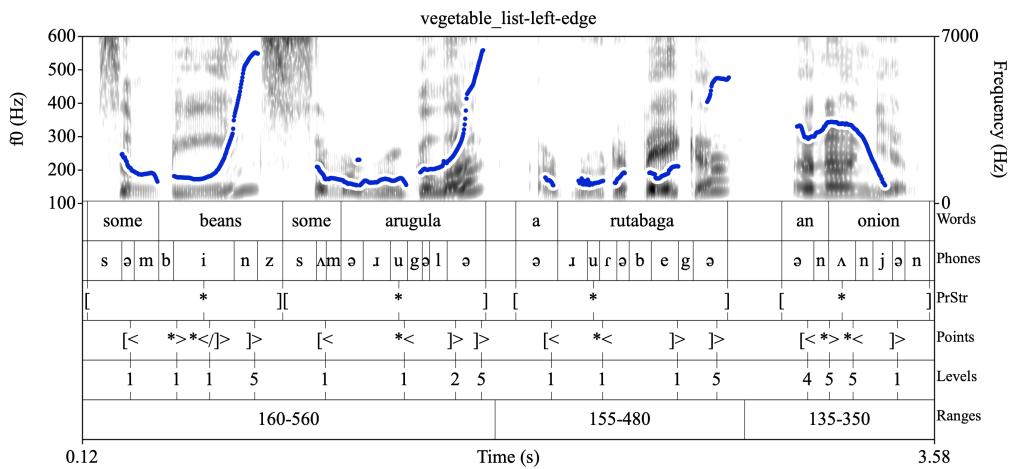
**Figure 1.18:** An initial high-f0 turning point is given a [< label in the Points tier, indicating affiliation with the [ object in the PrStr tier.

As with other ambiguous cases as described above, there may be times when the signal is ambiguous, so that where there is a high f0 at the beginning of a phrase, a labeller may be uncertain whether there is a prominence or not. In such cases, the labeller can capture this ambiguity with a label such as [</\*< (accompanying, for example a [ and ?\* in the PrStr tier.), as is shown in Figure 1.19. This label indicates that the high initial f0 may be associated with either the left boundary, the prominence, or both.



**Figure 1.19:** An initial high-f0 turning point is given the label [< /\*< label in the Points tier, indicating that the labeller considers the affiliation ambiguous between an initial phrase boundary and a potential prominence, labelled [ and ?\* (respectively) in the PrStr tier.

Researchers using PoLaR may opt to attribute any initial f0 points to phrase beginnings. The example `vegetable_list` from Section 1.2.1.1 is repeated here in Figure 1.20 with Advanced Points tier labels indicating that initial f0 points are being attributed to the beginnings of each phrase.



**Figure 1.20:** The example `vegetable_list` repeated from §1.2.1.1, with both phrase beginnings and ends annotated in the PrStr tier. Advanced Points tier labels here indicate that the labeller associates f0 points with phrase ends or beginnings.

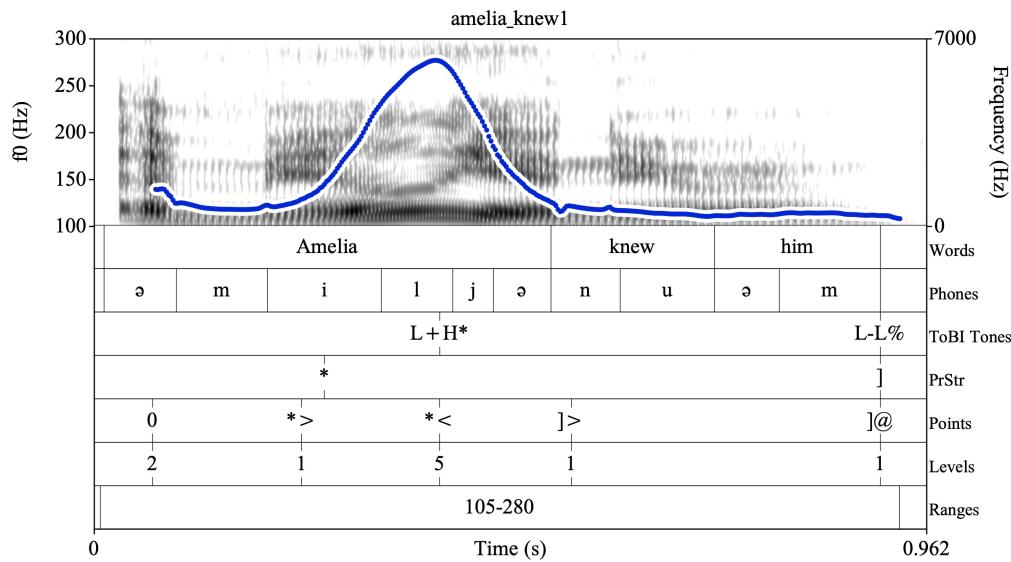
### 1.3.5 More Examples of Advanced Points Tier Labels

When using Advanced labels, it is important for prosodic labellers to be systematic so that data can be retrieved and aggregated with similar contexts or compared against contrasting

contexts. Here, we present a series of examples to help annotators establish and internalize the PoLaR system with respect to Points labelling.

The next two examples both contain prominence-related rises, and Example 2 also contains a boundary-related rise. Those labellers with some familiarity with English intonational phonology may have strong intuitions about which phonological event (on the PrStr tier) each of these rises is associated with. The Advanced PoLaR Points labels attempt to capture these intuitions.

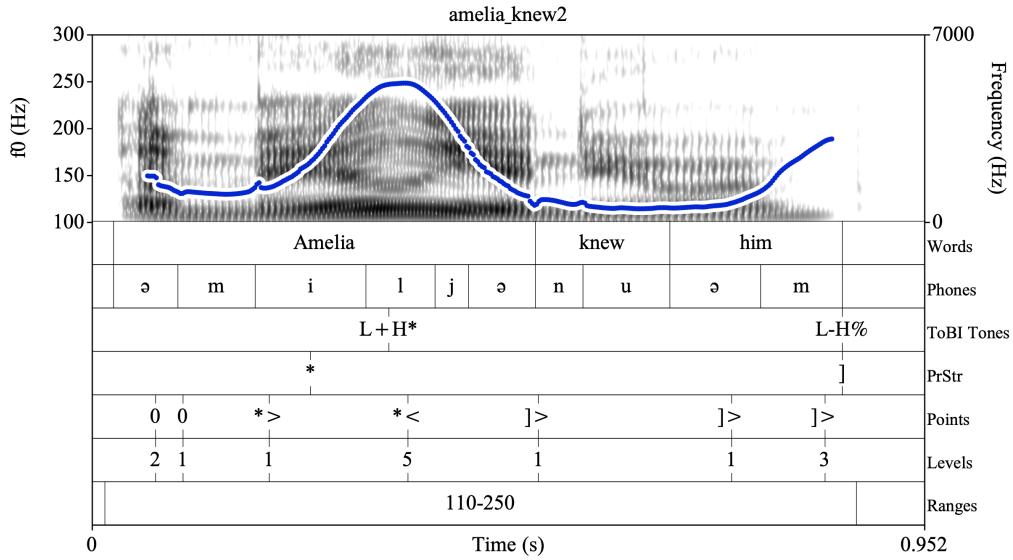
**Advanced Points Example 1: Points and PrStr** First, examine this example of turning points in the pitch track each associated with distinct objects in the Prosodic Structure tier:



**Figure 1.21:** amelia\_knew1, with Advanced PoLaR labels.

Again there is a single prominent word and the utterance is comprised of a single intonational phrase. In this case, the annotator has labelled the beginning of the reliable pitch with a '0', indicating they don't believe this initial pitch is related to any particular PrStr object. Next, the rise in pitch (described with two f0 turning points: one at the base of the rise and one at the peak) is analyzed as being associated with one phonological object: the prominence labelled on the PrStr tier (at the midpoint of [i] vowel in Amelia, as transcribed in the Phones tier). We annotate our analysis (that these two turning points are associated with this single phonological object) by using the '>' and '<' labels for two Points tier objects. Next, the pitch falls to a steady low, which has been analyzed as related to the phrase boundary labelled on the PrStr tier. The beginning of this steady low is annotated with a ']>' label. The end of this steady low has been annotated with a ']@' label: the ']' label on the PrStr tier and the f0 turning point on the Points tier are positioned at the same point in time, and so a @ pointer is used (instead of the more common '>' or '<').

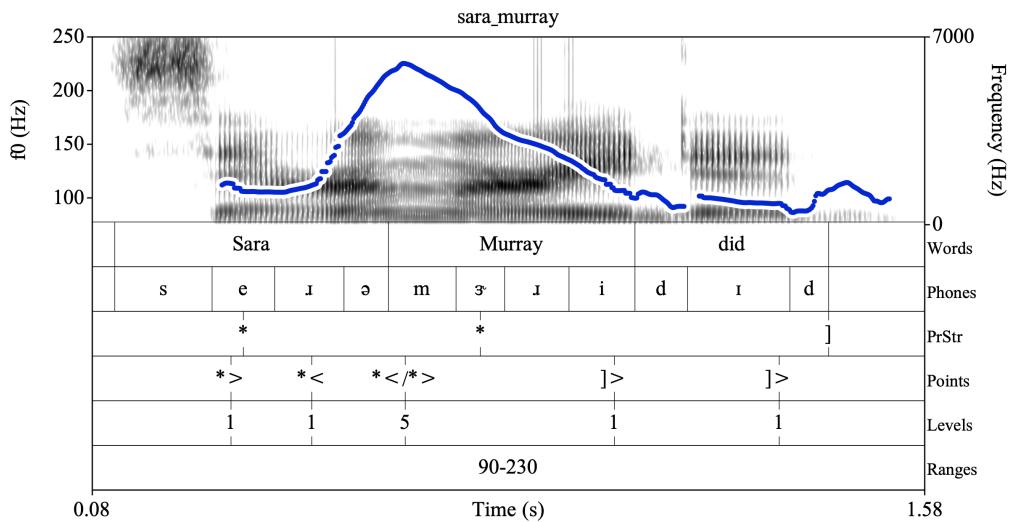
**Advanced Points Example 2: A minimal pair with above** This example differs minimally from the one above: there are two rises in pitch. The first is related to the prominence (as before), but there is a second rise related to the phrase boundary.



**Figure 1.22:** amelia\_knew2, with Advanced PoLaR labels.

In this example, three ']>' labels are used. The first two bookend the low plateau, and the third is for the endpoint of the final rise. By assigning all three of these movements ']>' labels, the labeller indicates that they analyze the low plateau and the final mid rise as all related to the phrase-final boundary. Thus, PoLaR provides a means for a labeller to distinguish boundary-related rises and prominence-related rises.

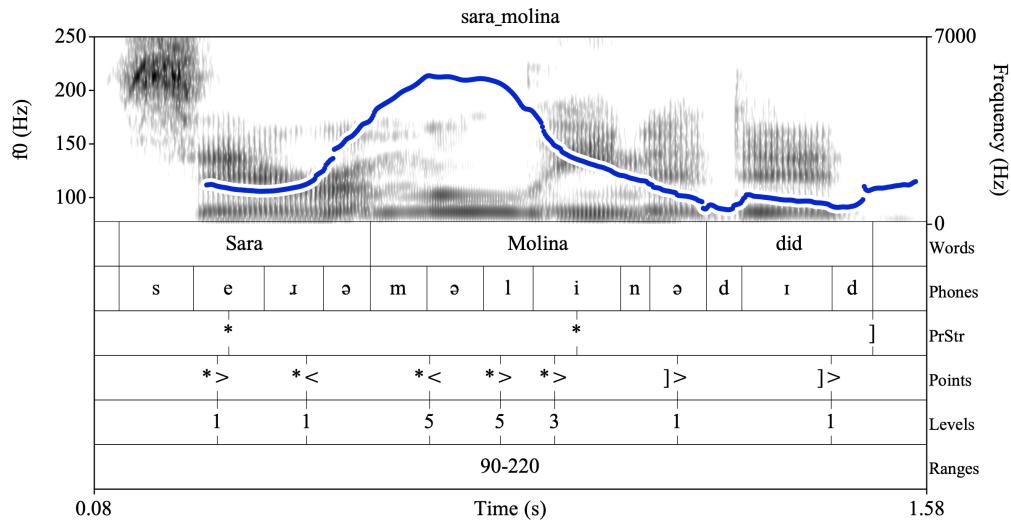
**Advanced Points Example 3: Ambiguous signal** As mentioned above, there are also cases where the signal might be ambiguous, such that a labeller could find support for an analysis where a rise may be a sort of ‘trailing tone’ for a previous pitch accent, or a ‘leading tone’ for a following pitch accent. An example of this is shown in Figure 1.23.



**Figure 1.23:** sara\_murray, with Advanced PoLaR labels.

As in the previous examples, most turning points can be analyzed as being associated with PrStr events. Turning points 1 and 2 are associated with the low pitch accent on *Sara* and the last two turning points are associated with the phrase boundary. However, the label on the 3rd point, ‘\*</\*>’, indicates that this turning point is possibly associated with the preceding pitch accent (‘\*<’; i.e., *Sara* is associated with a late-rising pitch accent), with the following pitch accent (‘\*>’; i.e., *Murray* is associated with an early-falling pitch accent), or possibly even both (as a kind of coalescence). Finally, the 4th and 5th turning points, both labelled ‘]>’, indicate that the labeller hears the word *did* as flat, with this plateau as cues to the phrase-final boundary. Before continuing on, one might note that there appears to be a visible shift in the slope of the fall curing the [.] of *Murray*. To see if this should be labelled as a turning point in the intonational contour, one can use straight-line approximation methods to attempt resolve the uncertainty (e.g., pitch manipulation, as discussed in section ?REF? of Chapter ?REF?). However, the results might not be definitive – in such a case, a labeller can use the ?0 label where they believe there may be a pitch turning point. (A labeller might also choose to postpone using the algorithm until later, and until this point is using ?0 as a kind of placeholder.)

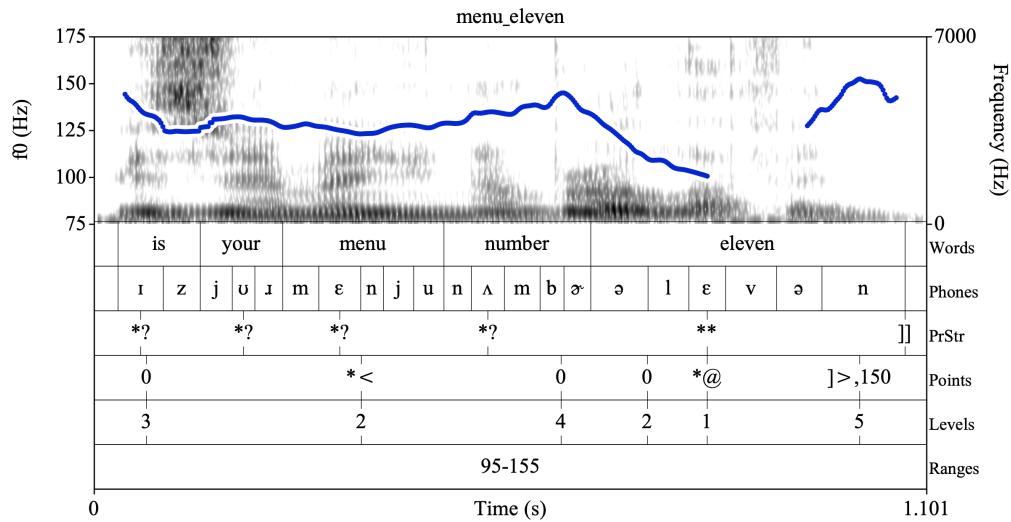
**Advanced Points Example 4: A minimal pair with above** In the previous example, *sarah\_murray*, there is a turning point that could be interpreted as cuing the \* on *Sara* or the \* on *Murray*. Compare that with what might be perceived as “the same” intonational contour, overlaid on a slightly longer utterance – one with an additional unstressed between the stressed syllables of *Sara* and *Molina*:



**Figure 1.24:** *sara\_molina*, with Advanced PoLaR labels.

In this case, we can see that there is a plateau stretching between the [m] and [l] of *Molina*. The labeller has indicated an analysis on the Points tier that says the prominence on *Sara* is a low-high rise (1-1-5, in the Levels labels) and *Molina* as a high-mid fall (5-3, in the Levels labels). Notably, instead of the ‘\*</\*>’ single point in *sarah\_murray* as in Figure 1.23, in this example there are two points ‘\*<’ and ‘\*>’ at the beginning of *Molina* (where there is ‘more space’ than in *Murray*). This supports the analysis that the ‘\*</\*>’ in *sarah\_murray* is a kind of coalescence, as suggested in the discussion above.

**Advanced Points Example 5: Non-PrStr-related Points labels** There is always a possibility that a labeller might want to analyze a movement (here, a rise) in pitch that is neither prominence-related nor boundary-related. Such an example is below:<sup>3</sup>



**Figure 1.25:** menu\_eleven, with Advanced PoLaR labels. Note the placement of the comma-override label at the end of “eleven”: it comes after the ‘]>’.

Since the initial f0 movements at the very beginning of the utterance are not always reliable, a ‘0’ here marks the first reliable f0. However, there is a perceptible rise, beginning with the low f0 (in *menu*). While the low f0 itself is here analyzed as contributing to a sense of prominence on *menu*, this labeller has used a ‘0’ label at the peak, to *avoid* analyzing the rise to the high in the unstressed syllable of *number* as a cue to a PrStr phonological object. In this way, while many A-M framework annotation systems do not provide labels for an f0 movement unrelated to prominence/boundary, PoLaR allows the labeller to indicate a phonological analysis of pitch turning points, but the labeller is *not required* to do so. Moreover, even when committing to a phonological analysis, these annotation conventions allow a labeller to systematically label f0 movements that are not predicted by the current state of their chosen phonological theory.

It is possible that a different labeller with a different set of theoretical tools and analyses might indeed want to label the turning point in *number* as related to some phonological object. PoLaR embraces this variety in approaches, and does not require that its labellers adhere to any particular theory. Instead, **what PoLaR aims to do is allow the labeller to explicitly annotate aspects of intonational phonetics** (e.g., in the Points tier), **intonational phonology** (e.g., in the Prosodic Structure tier), and **mapping relationships between the two** (e.g., with non-‘0’ labels on Points tier).

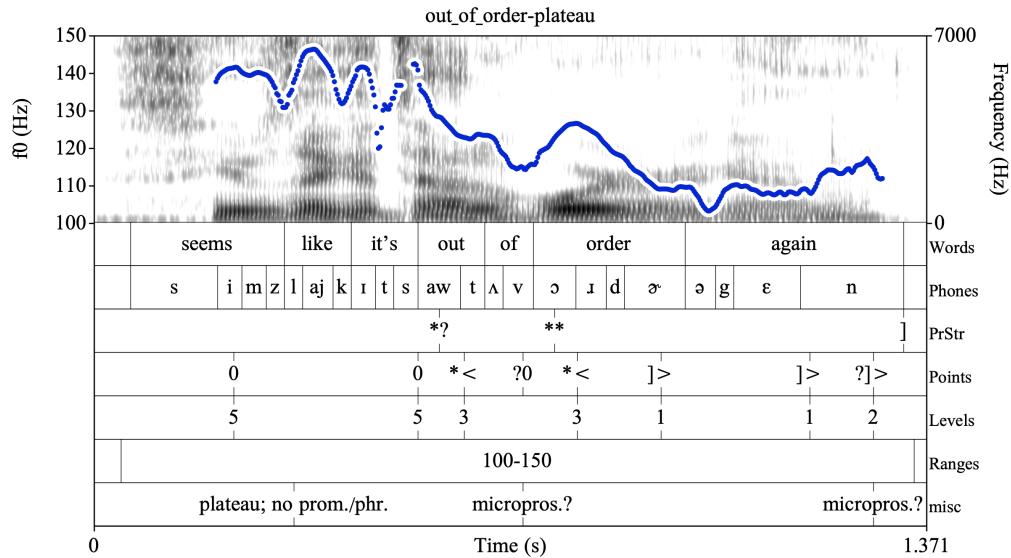
These non-‘0’ labels should be used whenever the labeller is confident in a particular phonological analysis.<sup>4</sup> On the other hand, if the labeller cannot decide on (or does not feel comfortable with) a (complete) phonological analysis, they can simply use the ‘0’ label for either or both of

<sup>3</sup>Note that the fourth Points tier label is ‘\*@’ and not ‘\*\*@’. The ‘\*@’, ‘\*<’, and ‘\*>’ labels are meant to point at any type of \* label (i.e., ‘\*’, ‘\*\*’, ‘?\*’) on the Prosodic Structure tier. Similar for ‘]>’ on the Points tier pointing to a ‘]]’ on the PrStr tier.

<sup>4</sup>Such labelled intuitions will be useful for later machine learning.

the two f0 turning points. As a general rule, the labeller is advised that they can always use a '**0**' label. To reiterate what was said earlier, the default should be seen as: every perceptually salient turning point labelled in the Points tier can be labelled with a '**0**'. Advanced labels are provided when the labeller is comfortable making assertions about the relationship between changes in f0 and intonational phonology.

**Advanced Points Example 6: No commitment to an analysis** Another example of using the '**0**' label for lack of phonological commitment is given below.<sup>5</sup>



**Figure 1.26:** *out\_of\_order-plateau*, with Advanced PoLaR labels.

We have already discussed the plateau at the beginning of this utterance, in Chapter [?REF?](#). Given the labellers' lack of phonological analysis for the plateau, these labels remain as '**0**'s, even when the label is otherwise using the wider array of Advanced Points labels.

This example also has two notable movements in the f0-tracking where the labeller is uncertain. The first is the dip in the tracking between "*out*" and "*order*", and the other is the rise in the tracking at the end of "*again*". In both cases, the labeller has begun their Points tier labels with a '?' to indicate their uncertainty about whether there is a turning point. (One factor contributing to the uncertainty is likely that this utterance is produced in a very narrow pitch range: 100-150Hz, making it hard to ascertain whether a small f0 movement is significant enough to merit a Points label or whether it may be a microprosodic effect, as in the case of the dip during "of" that might be related to the [v].)

<sup>5</sup>A possible MAE\_ToBI label for this would have an H\* on *seems* and an H+!H\* on *out*. Such labelling would indicate either (i) *seems* is prominent, or (ii) H\*s occur on non-prominent syllables. Indication (i) would not comport with the labellers' intuitions here, that *seems* is not prominent (see Prosodic Structure tier). While indication (ii) would require non-canonical views on MAE\_ToBI, some labellers may adopt it; but adopting such a view would make H\* a label for prominence and high pitch, or no prominence and high pitch. PoLaR gets around these complications by giving the labeller the freedom to explicitly label prominence separately from the perceived pitch.

### 1.3.6 Summary of the Advanced Labels for the Points Tier

The PoLaR annotation system, which is neither purely phonetic nor purely phonological, provides a means to make explicit how labelling of acoustic cues is phonologically informed. Moreover, PoLaR is flexible enough that it does not require its labellers to ascribe to any particular phonological model. (Even the set of labels in the Prosodic Structure Tier could be modified to fit different phonological models.) This allows the labeller to not only annotate phonological objects (in the Prosodic Structure tier) and phonetic cues (e.g., f0 turning points), but also to indicate a mapping relationship, all without requiring commitment to a particular phonological model (to avoid being an impediment towards wide adoption).

The table below lists 7 Points Tier objects that a PoLaR labeller can use. It is possible to add a ‘?’ before each label, to indicate labeller uncertainty about the presence of a pitch turning point, and it is also possible to include a numeric value after each label, to indicate labeller intuition about the pitch value (in Hz).

| <b>Label</b> | <b>Relevant Prosodic Structure object</b>      | <b>Relative to an object on the PrStr tier, this f0 turning point is time-aligned...</b> |
|--------------|--|--|
| <b>0</b>     | No phonological analysis given by the labeller | N/A  |
| *>           |  | ...before the relevant *   |
| *<           | Prominence: *, ?*, or **                       | ...after the relevant *  |
| *@           |  | ...with the relevant *   |
| ]>           |  | ...before the relevant ]   |
| ]<           | Phrase boundary: ], ?, ]]                      | ...after the relevant ]  |
| ]@           |  | ...with the relevant ]   |
| [ >          |  | ...before the relevant [   |
| [ <          | Phrase boundary: [ , ?[ , [ [                  | ...after the relevant [  |
| [ @          |  | ...with the relevant [   |

**Table 1.6:** An expanded set of labels for the Points tier, with several advanced options. The fundamental label ‘0’ is highlighted, with the remaining labels being for labellers with more experience. Note that, as always, a ‘?’ can be prepended to any of these labels to indicate labeller uncertainty about whether there is indeed a pitch turning point at this time.

(Note: an individual ‘]’ on the PrStr tier is very often associated with multiple ‘]>’, ‘]<’, or ‘]@’ targets. Similarly, this sort of many-to-one relationship is also very common for prominence-type annotations.)

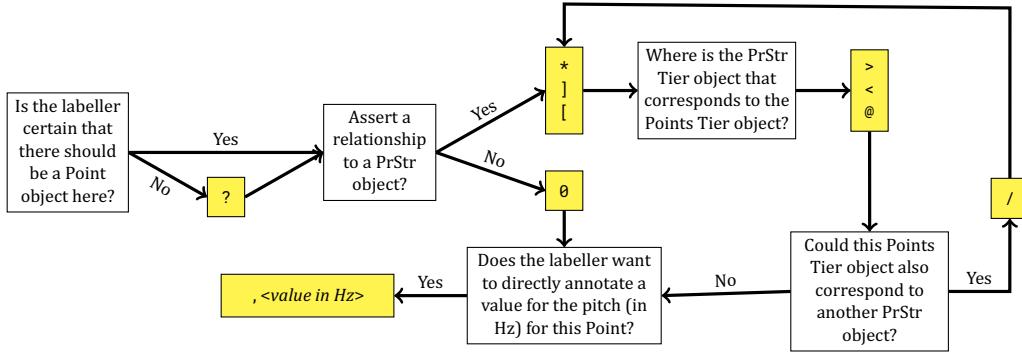


Figure 1.27: A finite state grammar for Advanced Points tier labels.

## 1.4 Ranges (Advanced)

As described in Chapter ?REF?, the Ranges tier defines the pitch space in which the Points tier labels are to be interpreted via the Levels tier, giving it a critical role in describing local pitch dependencies, such as within a prosodic phrase. But a range is not always limited to a single phrase; a further advantage of annotating ranges is the ability to annotate more complicated pitch relations across longer stretches of speech, allowing labellers to capture observed/intuited cross-phrase relations. By delinking the designation of pitch ranges from phrasing, PoLaR can shed light on the complex ways in which tonal cues interact with other cues to prosodic grouping and prominence. In this section, we discuss in detail some ways in which the Ranges tier labels can reflect labeller intuitions about relations among local ranges, as well as interactions across PoLaR tier labels. We will look at common types of range changes that occur within even a short utterance, as well as how ranges may interact across longer stretches of speech.

The idea of variable local pitch ranges within an utterance may be familiar to those labellers who have experience with other labelling frameworks, such as ToBI, in which it is generally understood that a local pitch range can change in the middle of the utterance. For example, the ToBI guidelines indicate that, even within a single utterance, “[e]ach new intermediate phrase represents a new paradigmatic choice of pitch range” (Beckman & Ayers 1997:p.25). A familiar example is the highly compressed pitch range often used for a parenthetical (see, e.g., Dehé & Wichmann 2010; we discuss parentheticals and cases similar to them in later sections of this chapter).

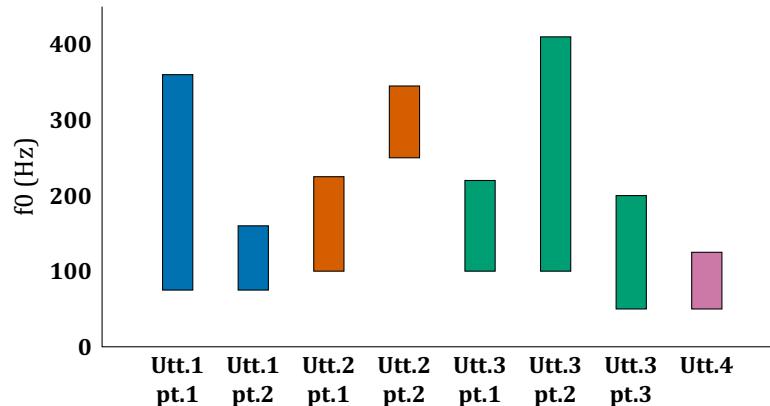
Even though there may be a connection between local pitch ranges and particular sorts of prosodic phrases, as made explicit in phonological models of intonation (as with ToBI; but see section 1.4.8, which contrasts PoLaR and ToBI treatments of ranges), PoLaR does not require the labeller to always attend to any particular definition of (syntactic/prosodic) phrase boundaries when making decisions about pitch range intervals. (i.e., A labeller need not (i) have a particular definition of constituents like “intermediate phrase”, (ii) be able to recognize constituents on the basis of phonetic cues, or (iii) have a theory of the alignment between constituent structure and pitch range changes.) Instead of relying on an existing linguistic theory that would predict where range changes will occur, the primary direction for labelling pitch range is that the labeller should rely on their intuition as to which regions of an utterance belong to separate pitch range intervals. (In the last part of this section, we will return to the idea of intuition, and what sorts of things a labeller may wish to attend to in sorting out their

intuitions.)

Because the ranges are labelled separately from prosodic phrasing, local pitch range intervals are not constrained by the boundaries. (This even allows Ranges and PrStr to be annotated separately, perhaps even by separate labellers.) Just as we saw that a pitch range could span more than a phrase (in the example, `tomorrow_morning`, in Figure ?REF? of Chapter ?REF?), we can also have ranges that begin/end in the middle of a prosodic phrase (as delineated by boundaries annotated on the PrStr tier). In other words, a single prosodic phrase may contain multiple Ranges intervals.

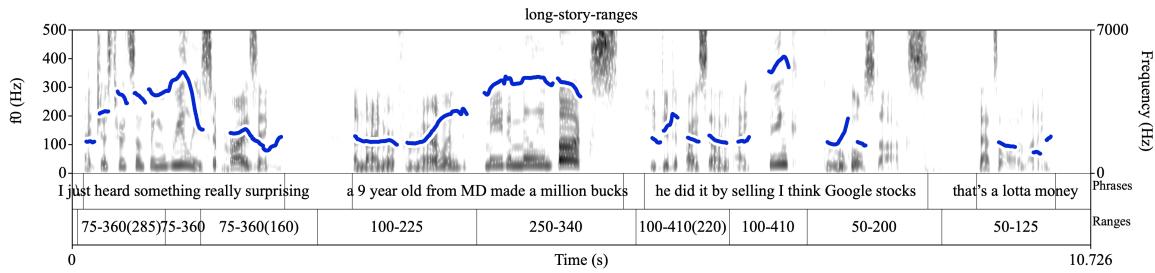
#### 1.4.1 A complex example of many range changes

Let us begin by imagining a set of utterances with a complex set of changes in ranges. A particular individual may have a physiological pitch range of 50-700 Hz, while their comfortable implementable pitch range may be 60-425Hz. As we have already discussed, the local pitch ranges will typically be smaller than either of these ranges, and may occur across various stretches of pitch space at various points during speech. Below is an idealization of four hypothetical utterances (in different colors) produced by a single speaker; note that some of these utterances are themselves composed of multiple local pitch ranges:



**Figure 1.28:** The y-axis is f0 (in Hz) and each of the bars indicates the speaker's local pitch range in particular (hypothetical) utterances; successive utterances are indicated by color changes.

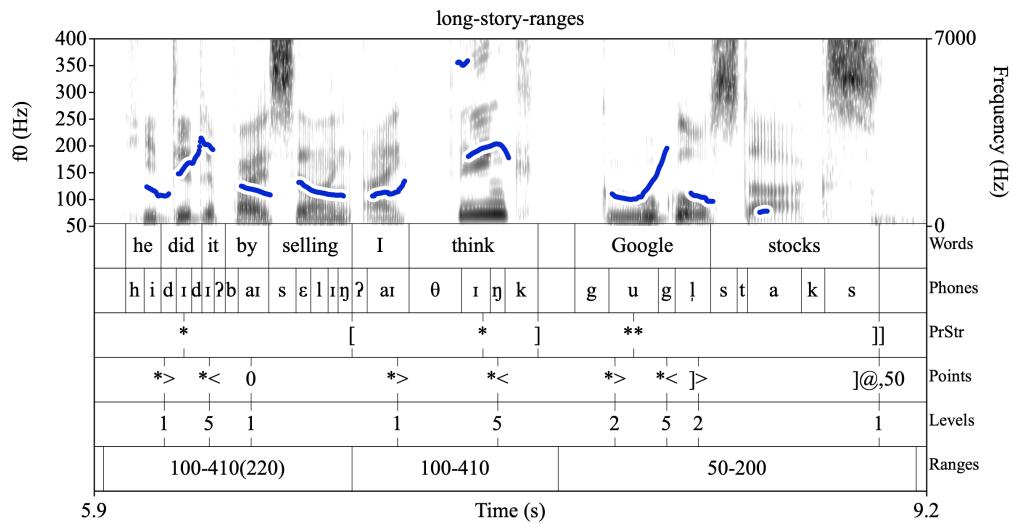
To reiterate, the Ranges that are annotated as intervals on the tier are intended to indicate what the labeller believes would constitute a “high” pitch in this local interval of time, and what would constitute a “low” pitch in the same interval. In this way, what the labeller is doing is making some postulations about how they believe the speaker is setting their pitch range during a particular time interval. An example that mirrors the idealization in Figure 1.28 is provided, in Figure 1.29.



**Figure 1.29:** An example of four utterances with nine pitch range intervals. Note that words are not aligned with the speech signal or the Range constituent intervals.

Consider the interval following the first pause (*a 9 year...*), labelled ‘100–225’, and compare that to the following interval (*made a...*) labelled ‘250–340’. The specific annotation indicates that the labeller perceives that a pitch of 225 Hz on *Maryland* would certainly count as a “high” (and would be transcribed with a value of ‘5’ on the Levels tier) in the first part of this utterance. However, moments later (during *made a...*), 250Hz is perceived as the “low” floor of this interval (and would be transcribed with a value of ‘1’ on the Levels tier).

In particular, the relevant aspect of intuition concerns “what goes together” as a pitch range constituent. To make this clear, let’s zoom in on *he did it by selling I think Google stocks* in the previous example and discuss why the labeller has labelled *I think* as a pitch range (sub)interval, rather than including *I* as part of the preceding pitch range interval, despite the fact that it begins with a low f0 very close to the f0 of the preceding word.



**Figure 1.30:** A closer look at some Ranges intervals.

In this figure, we can see that the three pitch range constituents map onto the following strings of words: *he did it by selling*, *I think*, and *Google stocks*. Looking at this pitch track alone, it isn’t obvious that *I* should be part of the same pitch range constituent as *think* – that is, one might wonder whether *think* should constitute a complete pitch range constituent in itself, such that *I* is part of the preceding one. There are a few reasons why the labellers in this case have grouped

*I think* together, all of which are strongly linked to their intuitions as English speakers. First and foremost: it is clear that the f0 during *think* is very high, in a different local range than the “high” pitches in surrounding words; while it is therefore clear (solely on the basis of measurements) that the local pitch maximum is around 410Hz, determining a value for the local pitch minimum requires consulting some intuitions about how words are grouped together and/or analyzing some additional cues to grouping in the signal. What are the relevant sources of information that underlie these (tacit) intuitions and (conscious) analyses? The answer lies in constituency (of various types), and the cues to constituency.

Let’s step back and look at the basic ideas behind local pitch ranges. We assume that local pitch ranges (which is what the Ranges tier aims to annotate) map onto constituents of some type(s), even if we are not entirely certain yet what those are. Language encodes various types of constituency, whose boundaries determine various linguistic properties: (i) ability to be manipulated syntactically, (ii) domains of semantic interpretation, and (iii) senses of juncture in speech informed by (for example) the beginnings/ends of coherent pitch contours, the local pitch range, placements of silence, final lengthening, and initial strengthening, among many other properties.<sup>6</sup> In this way, finding the beginning/end of a pitch range interval can be aided by finding other cues to the beginning/end of a constituent (with respect to any one of these linguistic properties; or more likely, with respect to multiple coinciding properties).

Returning to the example, let us review three reasons why one might want to label the Range interval constituent as the word sequence *I think* (and, e.g., not just the single word *think*). Let’s start with an f0-based intuition about constituency. When the annotators labelled this, they intuited that there is a single pitch movement —a large rise from low to high— which includes the pitch values during *I*. That is, the relevant pitch movement does not occur solely within the word *think* (where there is a rise from around 325Hz to 410Hz); instead, it is intuited as a rise from within *I* to within *think* (a rise of about 105Hz to 410Hz). It may be that if this interval consisted of continuous sonorant material, e.g. *I wink*, we would see more clear pitch-based evidence that there is a single f0 ‘scoop’ that starts in *I*. A second way in which prosody may contribute to the intuition of constituency might be based in the fact that there is a brief period of glottal closure between *selling* and *I*, resulting in a brief silence and initial strengthening in the form of a glottal release at the onset of *I*. Since glottal marking of this kind is common in American English (see Dilley et al. 1996), its presence here may contribute to the intuition that a new constituent begins at the onset of *I*. Third and finally, “*I think*” maps on to a syntactic/semantic constituent (the evidence for which we won’t discuss here). If we kept the f0 the same, but changed the words from ...*selling I think*... to ...*selling it but*... (where the extra high pitch falls entirely on *but*), there might be some cause to shift the Ranges interval to encapsulate only *but* – *it but* is not a syntactic/semantic constituent. (Of course, one would also need to weigh these factors against other cues, such as the placement of glottal events.)

To summarize the previous three paragraphs, there are three locally defined pitch ranges labelled above, and they may or may not map onto the boundaries of any currently-well-defined type of prosodic phrase (e.g., the ToBI intermediate phrase). In this way, instead of being forced to attend to predefined sets of cues, labellers are encouraged to attend to their intuitions (which are underwritten by a network of cues, not all of which are always present). In other words, the

<sup>6</sup>Not every boundary corresponds to all of these factors named above, and these cues can also be used to signal different aspects of structure (such as lengthening associated with prominence). Moreover, different types of constituents may be signalled by different types of cues. [It may even be that there are constituents which are only detectable through one of these properties, instead of a constellation of them.]

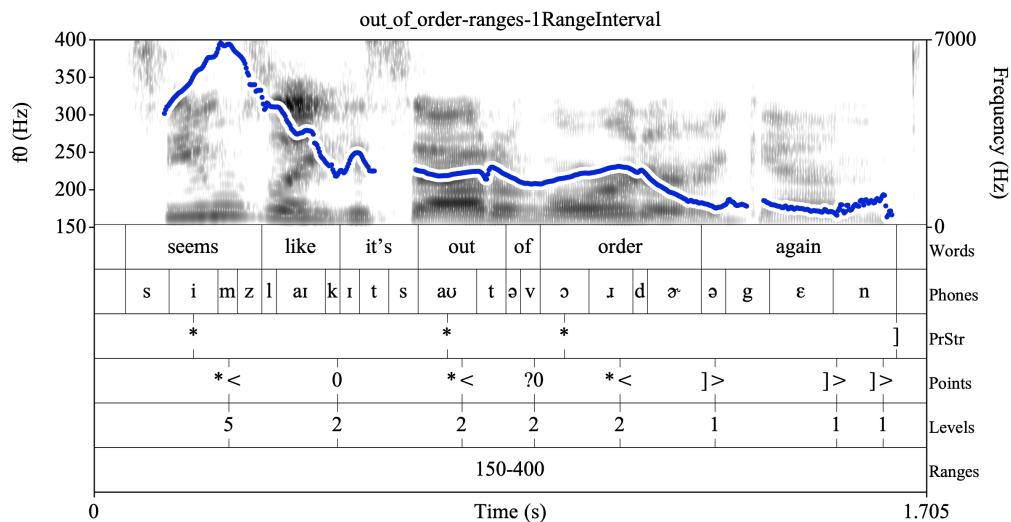
labeller should not feel restricted by particular perceptions about where (syntactic or prosodic) boundaries are.

### 1.4.2 How Many Range Intervals? Different Approaches and Their Implications

Whereas complete PoLaR labels must have at least one range annotated for every utterance, it will often be the case that an utterance has more than one range. A beginning labeller might start off by labelling ranges relatively sparsely, such as choosing a range for each interval of speech separated by noticeable silent intervals. As labellers become more comfortable with labelling, they may develop more detailed intuitions about local range changes, and therefore label more (and shorter) range domains for the same utterance that was annotated with a single range by a beginner labeller. That is to say, a more advanced labeller is likely to label more range changes than a beginner labeller.

One important goal of fine-grained labelling of ranges is that it allows for salient pitch changes from low to high to be captured using Levels annotations, even if those pitch ranges may be smaller than nearby range changes. (i.e., A fall that is perceived as going from high to low might involve a drop in pitch by 200Hz in one part of an utterance, where it may only involve a change of 50 Hz in another part.)

The recording in `out_of_order-ranges` exemplifies this issue well; in particular, both “*seems*” and “*order*” both sound high. In this passage, we will consider different ways of labelling the Ranges tier for this utterance. In Figure 1.31, the single phrase (note the lack of phrasing labels on the PrStr tier) has been labelled with a single Range interval (‘150-400’).

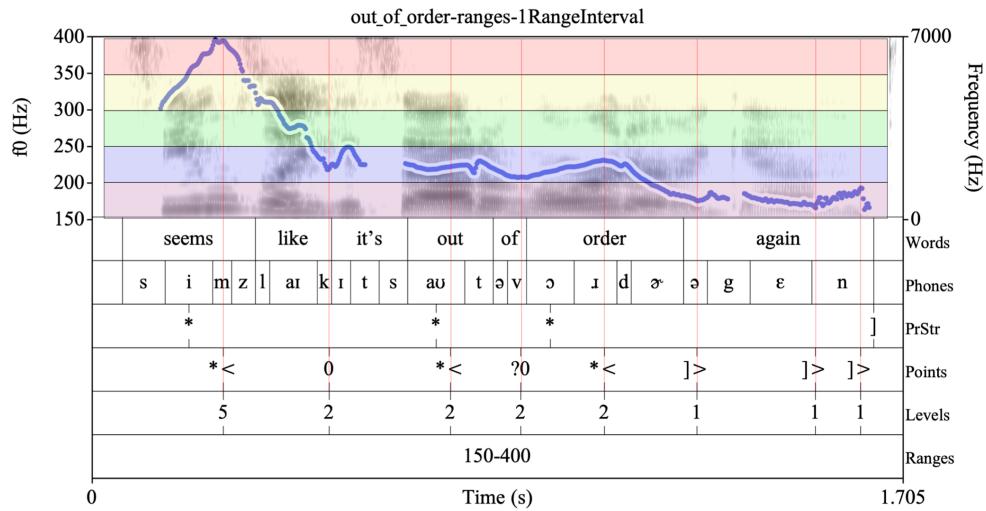


**Figure 1.31:** The recording in `out_of_order-ranges` labelled with a single Range interval for the entire utterance.

Whereas this approach to Ranges labelling is not wrong, per se, it does not capture many labellers' intuitions about the pitch contour in the second half of the recording. Namely, labelling only one long Range interval misses the intuition that “*order*” is perceived as high in the speaker's *local range*. (Note that this type of pitch shift downward, i.e. a local compression

of the range which includes what is commonly known as downstep, will be discussed in more detail in the next section.)

In addition, some small but still salient pitch movements are not reflected in the Levels, because of the (Basic) single range annotated in the Ranges tier. Listeners report hearing a pitch rise in the first syllable of “*order*”, but both Levels labels in this syllable have the same value of 2. This is because Levels directly follow Ranges labels, as illustrated in Figure 1.32, by way of the horizontal bands representing the pitch quintiles encoded as Levels labels.

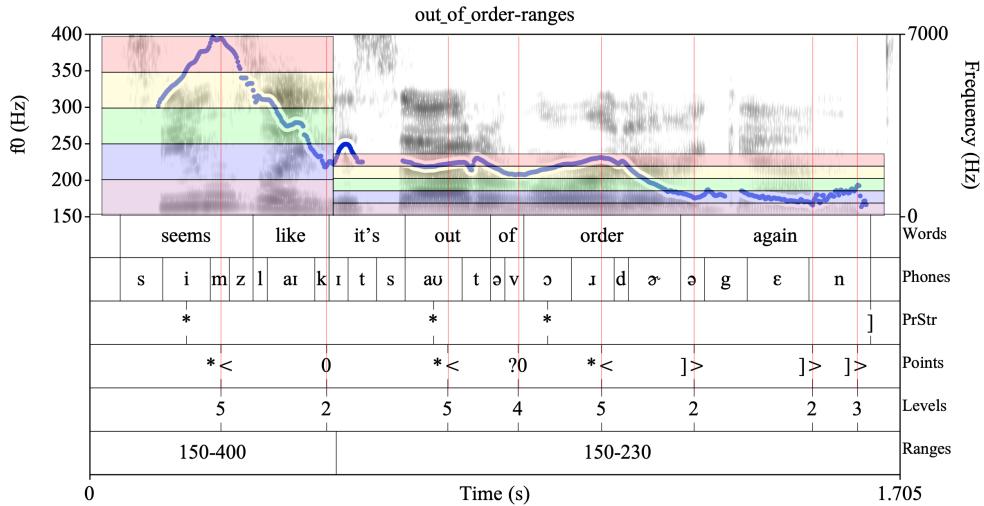


**Figure 1.32:** Each colored band represents the chunk of pitch space corresponding to each numbered Level, 1-5, as determined by the min and max values labelled in the sole Ranges interval.

In other words, a single Ranges interval here does not capture the intuition of a significant pitch rise during “*or*”, since that Ranges interval places both Levels labels in the same pitch quintile.

By contrast, to account for listeners who hear a rise in pitch during *order*, the Ranges tiers can be labelled as in Figure 1.32. The beginning of the utterance is given a range of 150-400Hz, but the second range has a compressed range of 150-230Hz, encoding a local rise in pitch.<sup>7</sup>

<sup>7</sup>Note that the first Ranges interval has a pitch floor of 150Hz, which is not attested in this portion of the utterance. We clarify why one might want this to be 150Hz in discussions of pitch floors (section 1.4.3) and Levels (section 1.5.1).

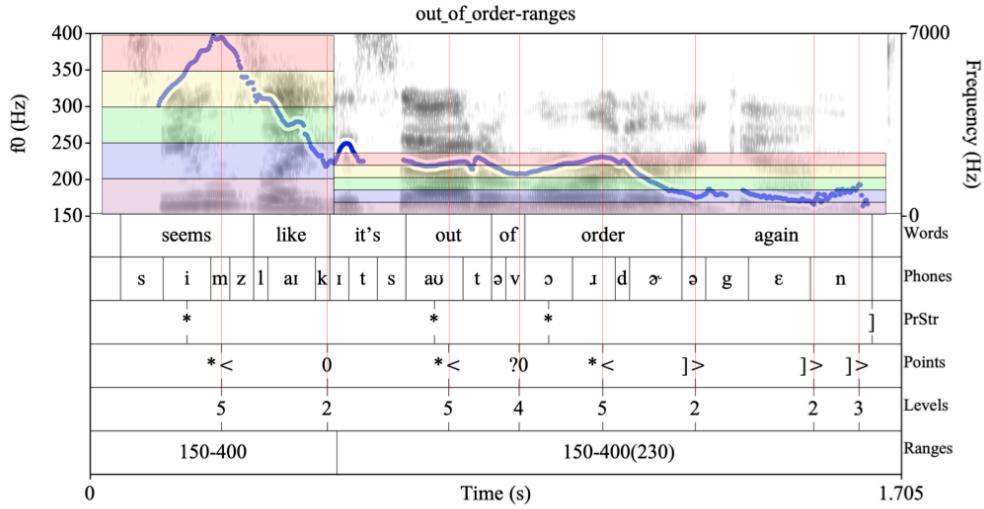


**Figure 1.33:** Each colored band represents the chunk of pitch space corresponding to each numbered Level, 1-5, as determined by the min and max values labelled in each of the two Ranges intervals.

The second Ranges interval is added because the labeller is annotating their perception/interpretation that “*seems*”, “*out*”, and “*order*” all sound high. The practical guidance is, when the pitch sounds high but looks low (relative to other pitch), the labeller ought to consider adding a new Ranges interval. We return to the issue of Levels labels and how they inform Ranges labels in section 1.5.1.

### Refining Basic Ranges into Range groups

Additionally, an advanced labeller who is dividing up Ranges intervals into smaller ones (as in Fig. 1.33) may want to preserve the more global aspects of the utterance, as indicated in the single, larger Ranges interval in Fig. 1.32. To do so, an advanced labeller can add optional “parentheses labels” to the labels of the floor or ceiling of a range (or both), when dividing a larger interval into smaller ones. Inside the parentheses, the labeller adds the more advanced intuition, while leaving the label outside the parentheses the same across the smaller intervals that are interrelated. This is illustrated in Figure 1.34, which shows the *out\_of\_order-ranges* example again, this time with two ranges and a parentheses label.



**Figure 1.34:** Parentheses labels are used to mark an advanced intuition that there is a second pitch range within what had been labelled as one pitch range by a labeller using PoLaR Basic labels.

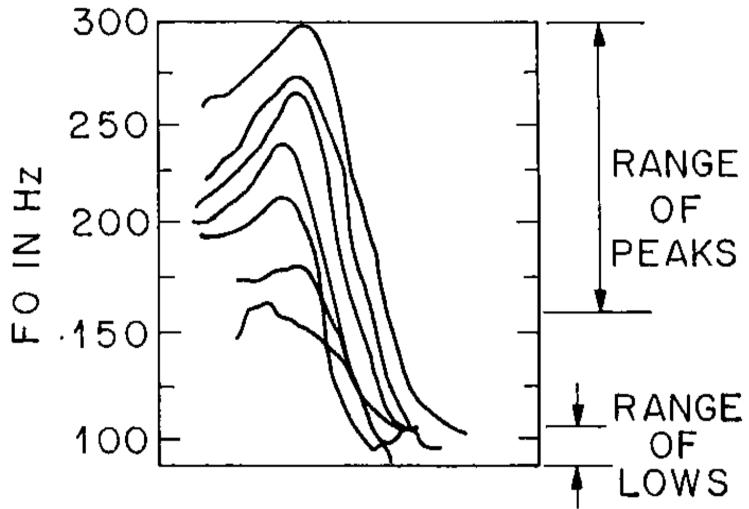
These labels could be generated stepwise as follows. First, a labeller adding Basic Ranges labels could label the entire recording as one Ranges interval of “150–400”; then, when an advanced labeller is adding more fine-grained labels, they notice the pitch range compression after “*seems like*” and so they divide the range in two, and in the second range they add a new, lower pitch ceiling label in parentheses: “150–400(230)”. In other words, here the value in parentheses indicates *the change* from the Basic label. Thus the advanced labeller has captured their intuition that the pitch range is compressed, in direct relation to the range labelled “150–400” (i.e., that this is some kind of range compression operation that subordinates the latter range to the former one).

In this way, parentheses labels in the Ranges tier can be used both in cases where there is a second round of labelling (first Basic, then Advanced, as in this example), and where an advanced labeller initially captures global aspects of groups of ranges. (See section 1.4.5 on subranges and superranges for more on this.)

In the coming subsections, we discuss other commonly occurring contexts in which one may be motivated to annotate multiple Range intervals for a single utterance or prosodic phrase.

### 1.4.3 Ranges and Pitch Floors

It has been observed that the pitch floor is typically more constant across long stretches, while the ceiling is much more variable: It is more likely that a labeller will find changes to the pitch range maximum, than they will to the pitch range minimum, especially within a particular utterance. (Though speakers do also make use of entirely different pitch ranges, with minima and maxima both differing, especially in cases of using falsetto voice or of intonationally distinguishing parenthetical speech, as discussed later in this section.) The observation that maxima vary more than minima, within speaker, is documented in Liberman & Pierrehumbert 1984, which provides the following figure (*ibid.*, p.159):



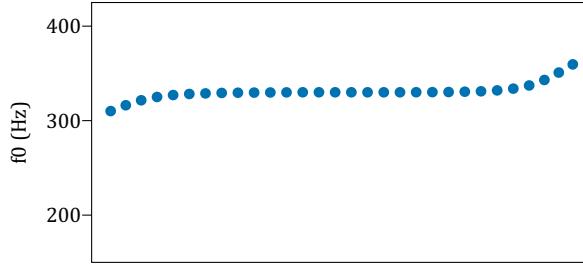
**Figure 1.35:** Pitch maxima seem to vary more than pitch minima, for a given speaker.(From Liberman & Pierrehumbert 1984)

Thus, in cases where determining a pitch minimum is difficult (such cases described in the preceding sections), a labeller may be able to draw inferences about a pitch range minimum by elsewhere in recordings by the same speaker. For example, one could compare with other utterances by the same speaker where they use a similar pitch range and the pitch minimum is more easily measured. (In fact, a labeller might not need to look further than other parts of the same recording to find a more easily measured pitch minimum.) Additionally, if a large enough set of recordings for that speaker exists, a labeller might be able to get a sense of the speaker's "normal" pitch range (its size, its floor, and/or its ceiling) and might be able to use this information to draw inferences about where the pitch minimum "should" be, even when difficult to measure directly.

#### 1.4.4 Some Trickier Cases with Local Pitch Ranges and Not-Attested Minimums

There are some fringe cases, which require deviation from the default approach, and where the labeller must attend to the more global context to identify the floor or ceiling. Not only will this deviation capture the labeller's intuitions but the automatic Levels Tier algorithm will produce more useful results (more on this in the next section).

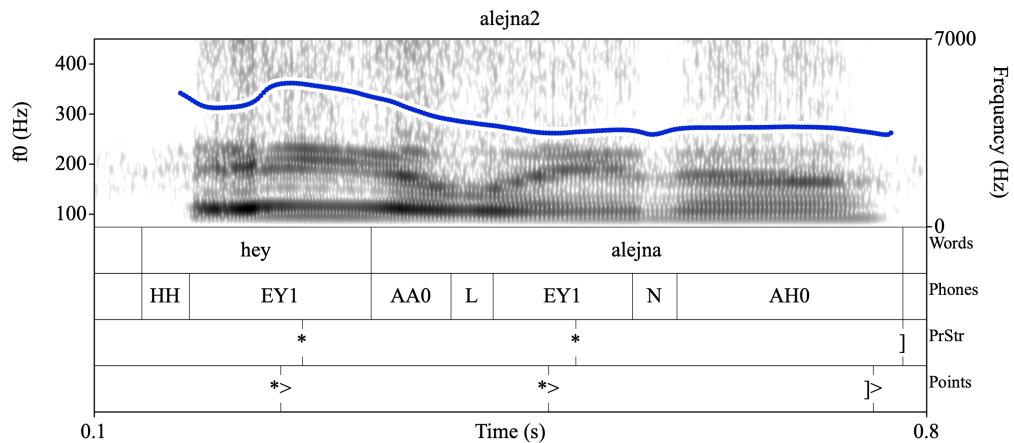
For example, the observed f0 minimum for a given interval does not always occur at the bottom of the local pitch range that the speaker gave themselves for a particular interval (i.e., the local minimum might not be a categorical "low"). This possibility raises an important question: how does an annotator know what to label as the mix/max for a local pitch range? As stated before in section ?REF?, the Basic Ranges default is to use the observed f0 points in the pitch contour as indicative of the minimum/maximum of the pitch range. However, you could imagine that a speaker may (in their mind) select a local pitch range but only ever produce observable pitch points in the high end of this range. As an example, an utterance may intonationally consist solely of high targets (e.g., a contour labelled in ToBI as /H\* H\* H-H%/). Such an example is provided in the figure below.



**Figure 1.36:** A hypothetical pitch track.

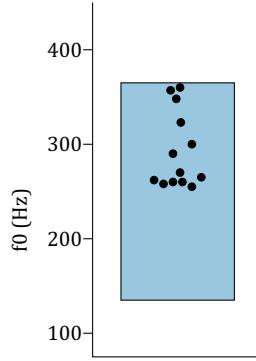
Assume that the labeller hears all these targets as high. As such, they have a good idea that the local pitch range max is something around 330Hz. At the same time, they see the local attested minimum value for the pitch is around 300Hz – labelling the Range as 300-330 would seem to miss the intuition that the minimum attested pitch sounds high.

In another case, a speaker may end up only producing pitch in the middle portion and high end of their selected local pitch range. For example, a speaker may start in the middle of this mental local pitch range, rise to high in that range, and fall to somewhere back to the middle of that range. An example of such a scenario is given Fig. 1.37, for alejna2 (we are assuming the speaker's comfortable pitch range is 75Hz-450):



**Figure 1.37:** A recording with an attested pitch minimum of about 250Hz to 365Hz.

On the basis of the  $f_0$  alone, the labeller would, by the default approach, label the interval as 250-365. However, it may be the case that the local pitch range (as represented in the mind of the speaker) is indeed something like 135Hz to 365Hz (meaning the speaker intended to never reach a “low” pitch – they fixed 135Hz as their pitch minimum, and only intended to reach a minimum of a “mid-level” pitch). This imagined scenario may be visualized as in the figure below, where each of the dots is an observable  $f_0$  point, and the box indicates the local pitch range for this utterance in the mind of the speaker. Note that none of the dots fall in the bottom portion of the range.



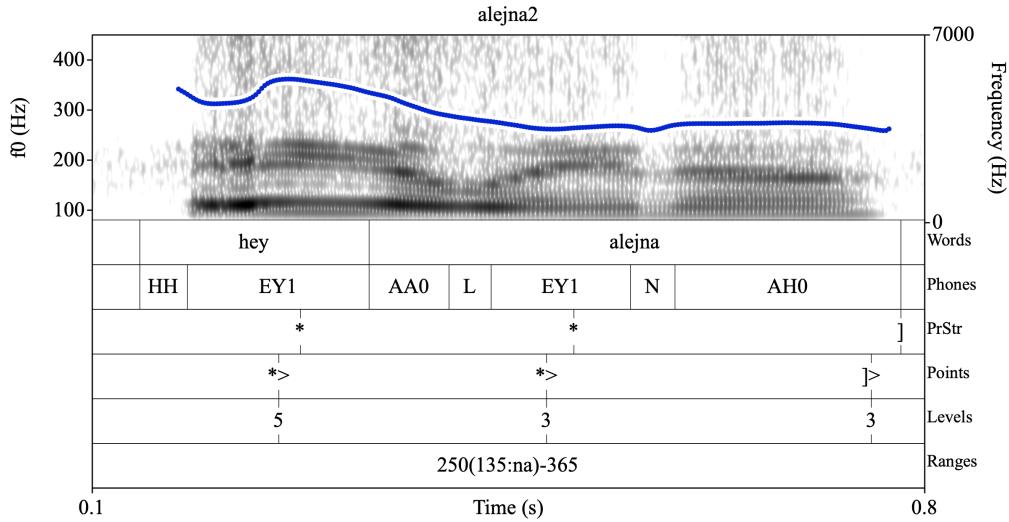
**Figure 1.38:** The filled box indicates the speaker’s intended local pitch range, and the black dots indicate f0 observations.

What this figure reflects is an intuition that the speaker did not reach the bottom of their local pitch range (i.e., that the speaker did not produce a low pitch in this utterance).

This would mean that in examples like Fig. 1.36 and Fig. 1.37 there are no local acoustic cues for determining even an approximation of the local pitch range minimum. This is because only a portion of the speaker’s local pitch range is actually attested by the f0 contour (e.g., the speaker is using only high, or only low, pitch targets). To be clear, this means there are cases where the labeller ought to deviate from the default practice of using observed f0 for deciding local pitch range max/min.

To allow for this, we provide a convention to capture that there are “unused” portions of the local pitch range, while also preserving what the attested values for the pitch range min/max are. In these cases, the labeller should label the pitch min/max with the attested value plus an educated guess for the “actual” floor/ceiling value plus “:na” (for “not attested”).

Solely on the basis of observed f0, a labeller would label the Range interval as “250-365”. Using this convention and matching the intuition described in Fig. 1.38, we add “(135:na)”, as below, to indicate that the pitch space extends down to 135Hz, even though no measured f0 is attested between 135Hz and 250Hz.



**Figure 1.39:** Labels capturing the intuition that the pitch range minimum (intuited as 135Hz) is not attested in the f0 track.

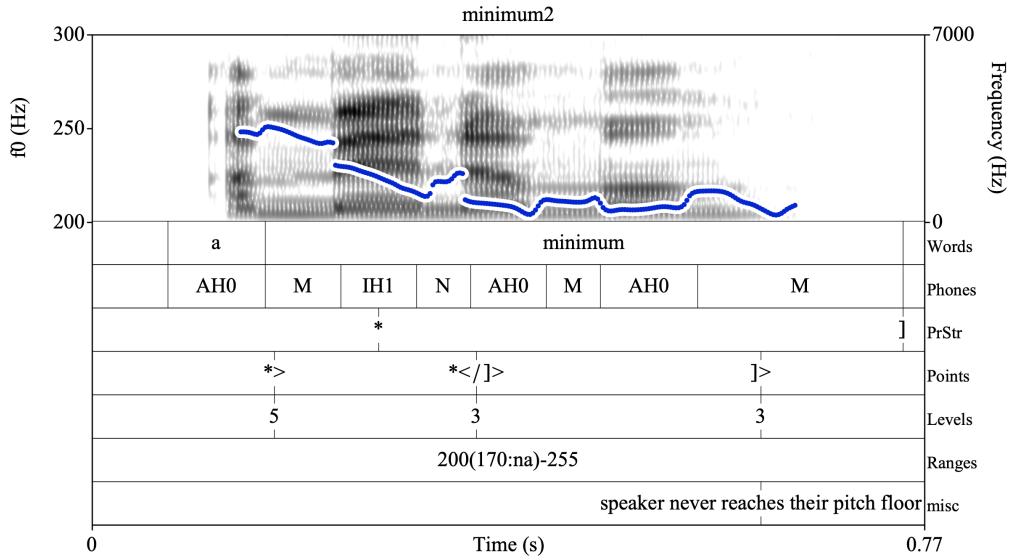
We must recognize that the value for the unattested minimum is meant to be an educated guess. We provide some guidance on how to best arrive at such a guess.

First, a labeller may know what this particular speaker sounds like when they produce a low – perhaps pulling from information about the speaker’s prosody in other cases (e.g., phonation, lengthening, or local pitch range size). Using this more global<sup>8</sup> information, a labeller may have the sense after listening to this particular recording that the pitch at 250Hz in this context is *not* the speaker’s low (even if 250Hz is this speaker’s low elsewhere, in other recordings).

Second, if a labeller hears the end of the utterance being in the middle of the pitch range, the labeller can do some simple arithmetic along the following reasoning: “250Hz is in the middle of the range, if the maximum is 365Hz, the minimum should be about equidistant from 250Hz – 135Hz.” Of course, a range of values would yield this same result of 250Hz being somewhere in the middle of the pitch range, so the labeller might just as well label this Range as “250(160:na)-365” or something else, if it aligns with other more global information better than 135Hz as the pitch range minimum. (We return in the next section to using Levels values as a sanity check on Ranges labels, in a similar way.)

Another similar example with this “:na” label is provided with `minimum2`.

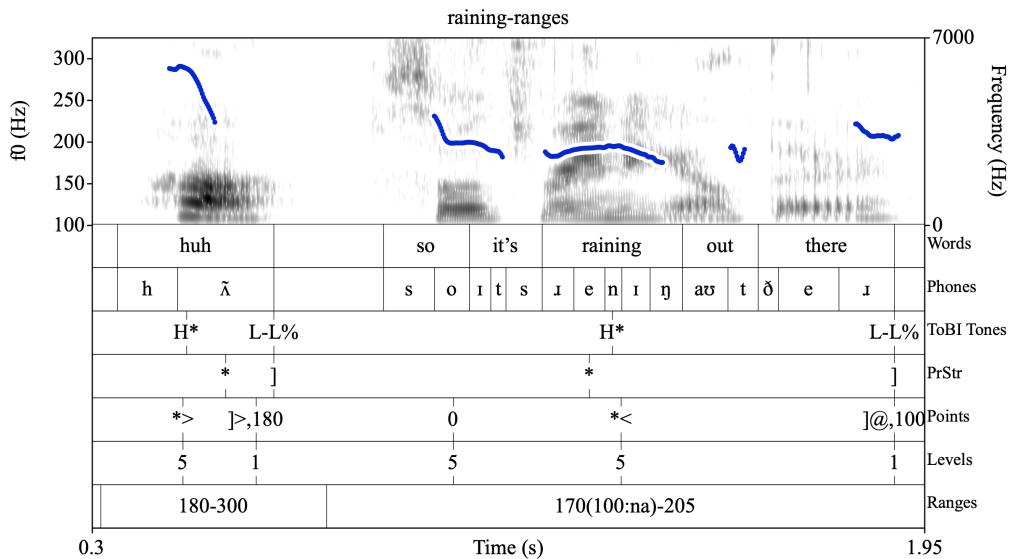
<sup>8</sup>In the data analysis, if there is enough data from the same speaker, it may be possible to reconstruct what the pitch min is in cases like this (e.g., on the basis of normal pitch range sizes).



**Figure 1.40:** Labels capturing the intuition that the pitch range minimum (intuited as 170Hz) is not attested in the f0 track.

Here the labellers hear the pitch as falling to a steady mid-level, and not to a steady low, and so the lower half of the pitch range must not be attested by the f0 contour.

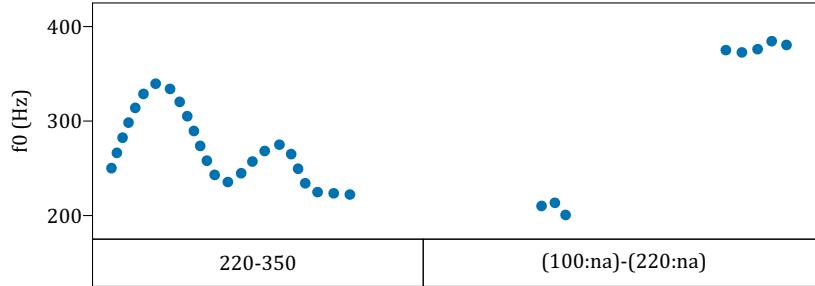
Another case where the labeller may want to use their ear is when the pitch track becomes untrackable or unreliable. For example, in an example like raining-ranges, labellers report being able to hear the pitch continuing to fall, even after “*raining*”, where the pitch track gets lost and/or becomes unreliable because of issues with voice quality. (i.e., Here the labeller knows that the end of “*there*” is not higher than “*raining*”, so the tracked pitch must be erroneous.)



**Figure 1.41:** The labeller provides an educated guess of the pitch minimum during “*out there*”.

Here the f0 measured at the end of “*there*” is measured as around 200Hz; as this might be an instance of pitch doubling, an educated guess for the final f0 may be 100Hz (as indicated by the comma-override), but the lowest reliable f0 (towards the end of “*raining*”) is around 170Hz. This leads to the Ranges label of “170(100:na)-205”.

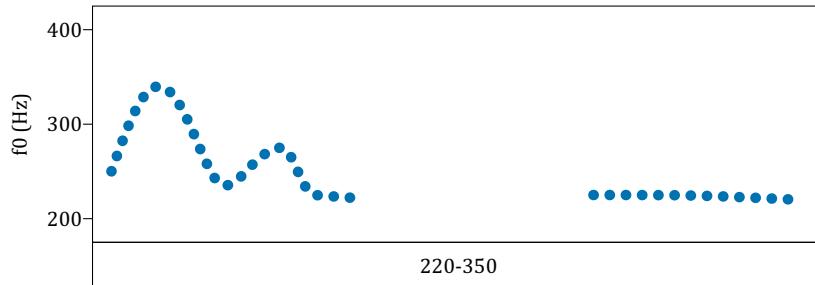
Finally, there are cases where there is what seems to be an entire Range interval where there is no (reliable) f0 tracking (e.g., because of low recording quality, or non-modal phonation). To be clear, this is intended to identify cases where the labeller has no visual cues to the pitch values, perhaps because the software-generated f0 track is missing or is tracked at the wrong pitch height (as determined by the labeller’s ear) for a long stretch of time, as in the figure below:



**Figure 1.42:** A hypothetical pitch track.

Here, the labeller thinks they are hearing the f0 as all lower than 220Hz (the previous range minimum), and is providing an educated guess (perhaps based on other recordings by this speaker) that the minimum is at 100Hz. However, the labeller does not want to commit to saying there are any attested f0 values, and so they have given a label where there are no pitch range values outside of the parentheses. Thus in all cases, labellers ought to give Ranges labels – even if made up entirely of guesses for which the f0 min/max are not attested directly in the f0 contour.

As a contrast with Fig. [fig:ranges what if 4], consider Fig. 1.43, where the f0 track becomes unreliable but it *does* recover (and the labeller hears it as the same local pitch range throughout), there is no need to add any “:na” labels (assuming the labeller does not hear any dip in pitch during the interval where no f0 is measured by the software).



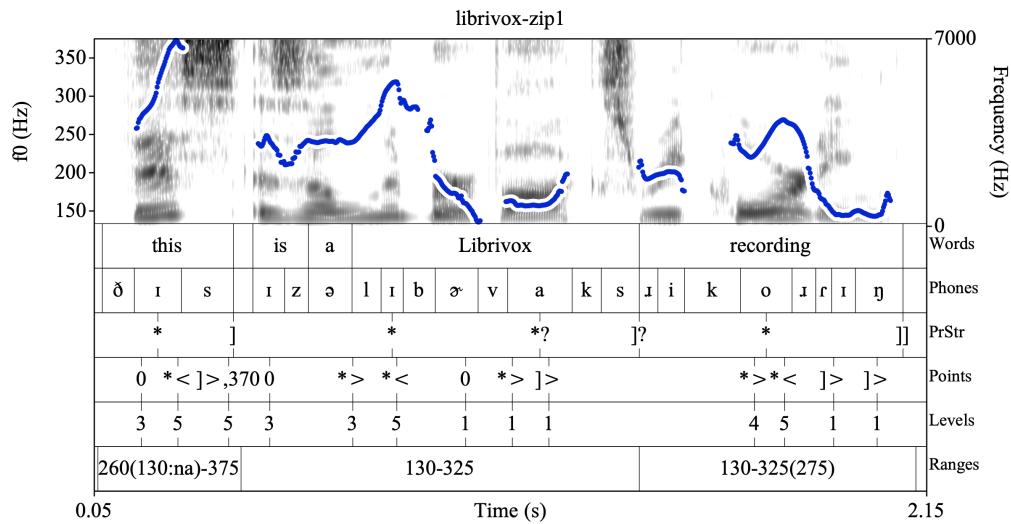
**Figure 1.43:** A hypothetical pitch track.

#### 1.4.5 Sub-ranges and super-ranges: Grouping together ranges using a common floor

In labelling the ranges of longer stretches of speech containing multiple ranges, the labeller may have the intuition that two or more nearby ranges (often consecutive, but not necessarily

so) are related tonally, and form part of a larger group. For labellers who have such intuitions and wish to capture them, they are encouraged to indicate this relationship by labelling related ranges with the same Ranges minimum value.

The figure below shows an example of a speaker producing a sequence of phrases produced with distinct pitch ceilings. The first range covers quite a short duration, and contains a single word, as well as just one phrase and prominence as marked in the PrStr tier. If a labeller were to only see this part of the utterance, it might seem like the floor should be around the lowest point of this particular group, i.e., the beginning of the utterance, or around 260Hz. This would suggest that the pitch movement, at the beginning of the word “*this*” goes from the local low to the local high. However, when the rest of the utterance is taken into consideration, the labeller’s intuition might be that the three ranges form part of the same larger tonal group and this can be captured, as it is here, by giving all intervals the same pitch floor of 130Hz. Further, this common-floor approach captures the sense that the word “*this*” does not sound as though it starts at a real local low in the speaker’s range – but rather somewhere in the middle of their range. By choosing the floor from the observed pitch track in the subsequent ranges, the labeller captures both the intuitions about the floor of the first range, but also that the three ranges form part of the same tune or melody, i.e., some larger tonal group.



**Figure 1.44:** librivox-zip1, with Advanced PoLaR labels.

This example further illustrates varying approaches to labelling the Ranges tier, and their effects. For instance, one labeller might consider that there is only one local pitch range for the whole short utterance, such as 130-375. Another labeller might want to capture three ranges, one for each of the three prosodic phrases, where both the ceiling and floor are determined locally, on the basis of what the listener can observe/intuit about pitch minima and maxima in the time span. A third approach (and the one used to label Figure 1.44) could be considered a hybrid of these approaches: the pitch maxima are locally determined based on observations/intuition, while the use of a ‘shared’ pitch minimum captures the intuition that the phrases form part of some hierarchically superior tonal group. It is left to individual labellers to determine whether they want to capture the granularity of ranges in an utterance and the relationships between them. As we will see below, these choices can impact the automatic Levels Tier labels.

### 1.4.6 Summary of the Advanced Labels for the Ranges Tier

| Label            | Phonological Object  | Example  |
|------------------|--|--|
| [min]-[max]      | [min] is the local pitch minimum, in Hertz, rounded down to a nearby number ending in a 5 or a 0; [max] is the local pitch maximum, in Hertz, rounded up to a nearby number ending in a 5 or a 0               | If the local pitch min is 244.2Hz and the local pitch maximum is 381.5Hz, the label should be '240-385'                                  |
| [min][[min2]:na) | [min] is the Basic ranges value for the minimum, and [min2] (written in parentheses) is the Advanced ranges label, capturing labeller intuitions for a value that is not attested locally within that interval | If the local attestable pitch min is 112, but the labeller intuits that the speaker has deliberately not reached a low '110(90:na)-385'  |
| [max][[max2]:na) | [max] is the Basic ranges value for the minimum, and [max2] (written in parentheses) is the Advanced ranges label, capturing labeller intuitions for a value that is not attested locally within that interval | If the local attestable pitch min is 192, but the labeller intuits that the speaker has deliberately not reached a high '90-192(300:na)' |
| X-[max]          | "X" represents a pitch minimum that the labeller cannot determine; [max] is the local pitch maximum, in Hertz, rounded up to a nearby number ending in a 5 or a 0  | If the local pitch max is 264.8Hz and the local pitch maximum is not easy to determine, the label should be 'X-270'                      |
| [min]-X          | [min] is the local pitch minimum, in Hertz, rounded up to a nearby number ending in a 5 or a 0; "X" represents a pitch maximum that the labeller cannot determine  | If the local pitch min is 138.2Hz and the local pitch maximum is not easy to determine, the label should be '135-X'                      |
| NA               | "NA" indicates a stretch of unreliable pitch tracking, in places where surrounding information cannot be used to infer the pitch range minimum/maximum ( <i>To be used in last resort scenarios</i> )          | If the pitch is a long stretch of unreliable pitch tracking, the label for that unreliable stretch is 'NA'                               |

**Table 1.7:** An expanded set of labels for the Ranges tier. The fundamental label is highlighted, with the remaining labels being for recordings with less clear pitch tracking.

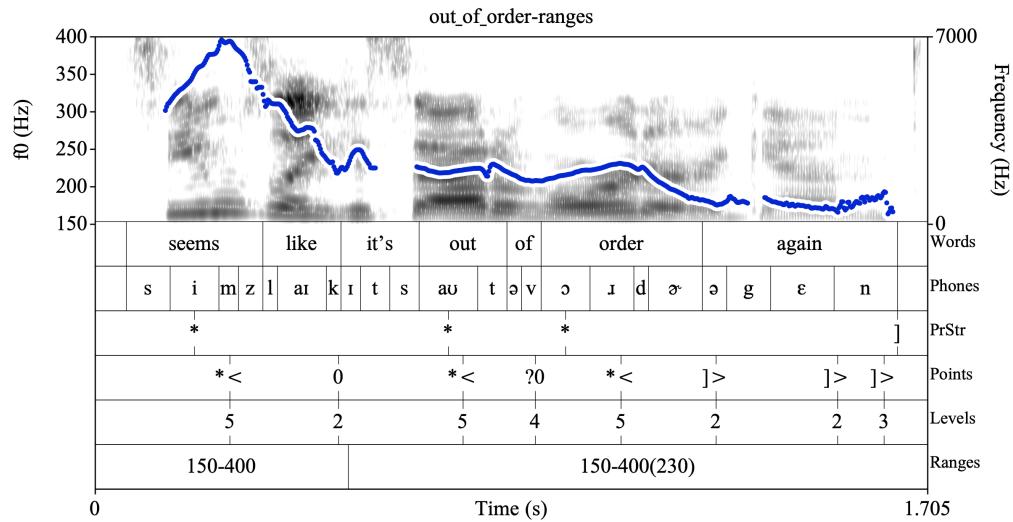
### 1.4.7 Some English Pitch Range Phenomena

In this section, we discuss some phenomena whereby pitch ranges are “shifted”, “expanded”, or “compressed”. In particular, we focus on examples from American English, and propose some Advanced labels to capture these phenomena.

#### 1.4.7.1 Range Compression: a commonly-occurring local range change

A commonly occurring context which corresponds to a change in local pitch range is when a tonal event is perceived as locally high in pitch, but is lower than a preceding high tonal event, which we will describe as “range compression”. A particular type of range compression in-

cludes what is commonly known as intonational “downstep”, which is where a perceptually high tonal target occurs at a (significantly) lower pitch than a preceding high tonal target. An example of this was given earlier, and is repeated in Figure 1.45.

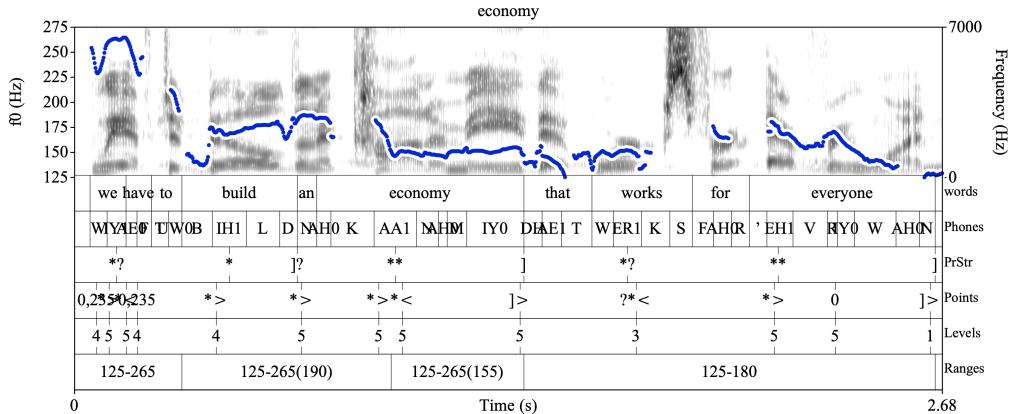


**Figure 1.45:** One prosodic phrase (PrStr annotation) and two local pitch ranges (Ranges annotation).

Here, *seems like* occurs in a pitch range with a much higher f0 maximum than *it's out of order again*, but the labeller also perceives *out* and *order* as high pitched. Though the pitch range changes after *seems like*, the labeller does not perceive enough of the cues to disjuncture (e.g., pauses, final lengthening, initial strengthening, etc.) for them to label different prosodic phrases on the PrStr tier. As such, the labeller does not annotate a **]** between *like* and *it's*, but does annotate a new Ranges boundary with a lower ceiling, to indicate that they hear *out* and *order* as high accents.

In this way, though some analyses would treat this drop in the pitch ceiling as due to the presence of certain phonological objects (e.g., a “downstep accent”), PoLaR treats all pitch range compression the same — with a new Range interval. In other words, the time-alignment of Range intervals is independent of prosodic phrase boundaries.

Thus, the main characterization of the range compression in this “downstep” context is the beginning of a new Range interval, the beginning of which is time-aligned to be where the pitch ceiling has finished falling. Another example exhibiting range compression is given in Figure 1.46, where there are two successive instances of range compression, and where the timing of the Range intervals is instructive.



**Figure 1.46:** Multiple Ranges intervals, where the labeller's analysis includes multiple instances of range compression.

In this example, the string *we have to build an economy* consists of a single prosodic phrase, and a novice labeller may label this entire phrase as a single Range interval of 125-265. A more advanced labeller (especially one trained to look out for downstepping) would notice that the pitch range is compressed twice, to account for the interpretation that *we*, *build*, and *economy* are all prominent and have high accents. The pitch ceiling falls from around 265Hz during *we* to about 190Hz during *build an*; the ceiling then falls again to around 155Hz in *economy*. What's especially instructive here is the timing of the pitch ceiling's fall during *economy* (which ToBI labellers may recognize as a H+!H\*): the pitch ceiling is dropping *during* the [a] vowel, and reaches the end of its fall about one third of the way into it. Since the Ranges interval labels track the pitch floor and the pitch ceiling, the new Range should begin where the change in range is completed; i.e., where the pitch ceiling's fall is completed. (Note that after *economy*, there is a new Range interval, whose beginning is time-aligned to the start of a new prosodic phrase. While ranges and phrases are independent, phrase boundaries do frequently coincide with new Ranges intervals, since phrase boundaries may trigger a pitch reset; for recent work on the coincidence of pitch reset and phrase boundaries, see Brugos 2015 and Kim 2020.)

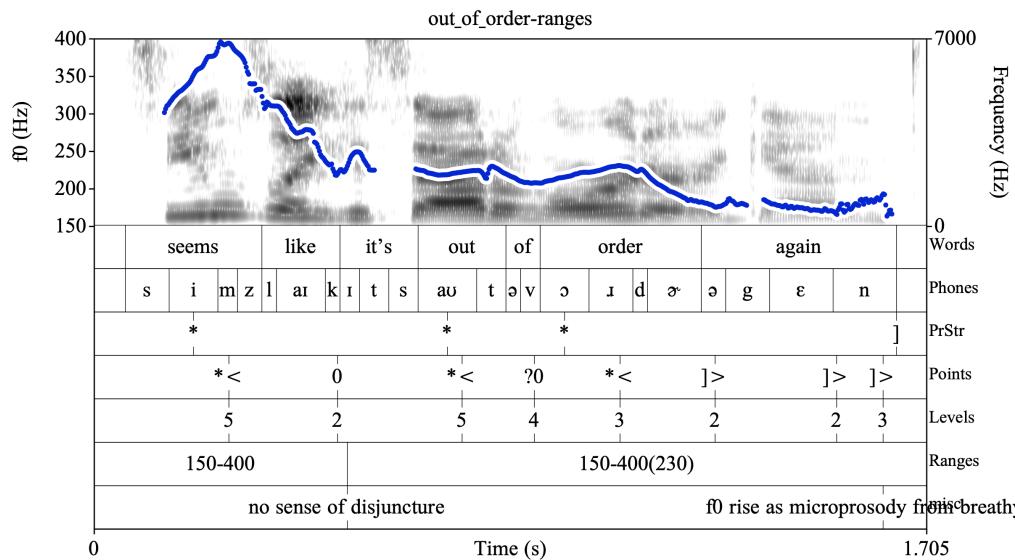
While the guidance is relatively clear in these cases of compression, Ranges interval labellers may at times find it challenging to determine how many range changes are appropriate, after several instances of compression, when the speaker's local range(s) gets very narrow. In such cases of ambiguity, as with Points tier labelling, we recommend erring on the side of over-labelling, and putting in more Ranges intervals rather than fewer.

It should also be noted that, while the labels in Figure 1.46 indicate an analysis that there are two instances of range compression (indicated by the usage of two subranges), this is an *analytical choice*, and equally valid (but analytically different) labels might not include one or both of these compressions. For example, under a different analysis, a labeller may treat the f0 fall during “*economy*” as a fall-to-low – if so, “*an economy*” should fall within the same interval on the Ranges tier. To be clear, this means that (like other Advanced PoLaR labels), different analyses can be tracked by using different Advanced Ranges labels. (PoLaR does not force any analytical choice; the usefulness of PoLaR is that it can keep track of whatever analysis the labeller prefers.)

### 1.4.7.2 Expanded ranges: when a range is compressed relative to what follows

Another type of local range change, one parallel to range compression as discussed above, is where subsequent prominences are realized with relatively higher pitch, which is sometimes referred to as “upstep.” PoLaR annotation can capture such pitch relations via the range tier in analogous ways to range compression relative to earlier highs.

The example another-little-banana shows such a case. Here, two prominences on *another* and *little* are both realized with high-sounding pitch, but the second prominence is noticeably higher. The labeller has captured that the f0 is high on both *another* and *little* by labelling two ranges, the first with a maximum of 245 Hz, and the second with a higher maximum of 330 Hz. (Note that these ranges are both within the domain of a single phrase in the PrStr tier.) Here, the Advanced ranges labelling captures the pitch expansion by means of labelling a lower ceiling in parentheses in the first range, reflecting that a Basic Ranges tier label would capture only one range with the higher ceiling for the whole utterance.



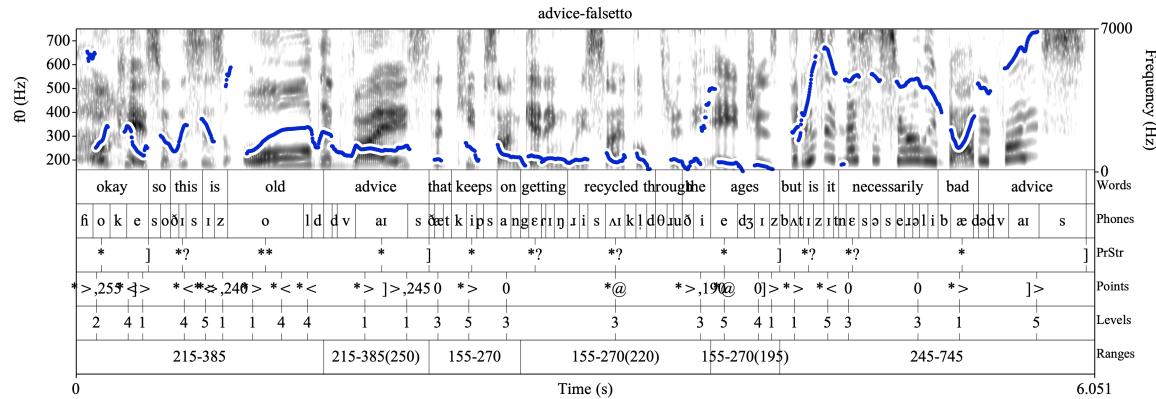
**Figure 1.47:** another-little-banana, with Advanced PoLaR labels.

Ultimately the motivation for these two range intervals here is to indicate that the labeller hears the pitch movement in *another* to be rising from low to high – and the Range tier annotations indicate the pitch floors and pitch ceilings. If only a single ranger were used, the Levels labels for *another* would not be 1-5, but 1-3 — and the 1-3 rise might not accord with the labeller’s intuition that the pitch during *another* reaches a “high”. This follows from the general idea that what Levels labels indicate is where “high” and “low” are (with the assumption that “high” corresponds to 4 or 5, and low “1” or “2”).

### 1.4.7.3 A few examples of range changes signalling parentheticals

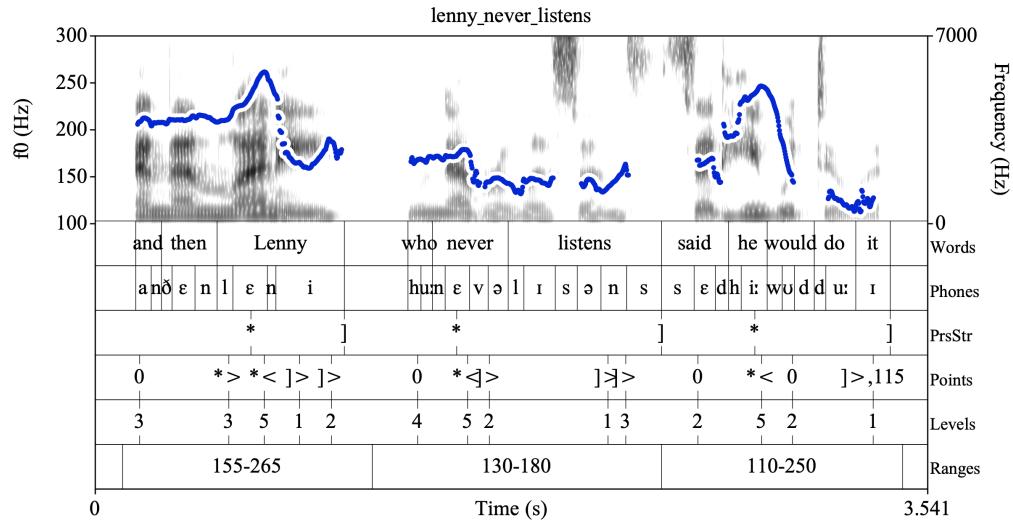
As mentioned in the section above on the relative stability of the pitch floor, there will be cases where both floor and ceiling change across ranges. One such example of this was already shown in the introduction to the Ranges tier in the Basic labels chapter, and is repeated here – note that this is not a parenthetical, but exemplifies a similar intonational phenomenon. In the third

range of the example advice-falsetto-NPR, the speaker has shifted the top of her range up to 750 Hz, and also shifted her local lows to a relatively high 250 Hz. In this example, the ranges shift at the locations of the phrase boundaries, as marked in the PrStr tier.



**Figure 1.48:** An example of both the pitch floor and ceiling of a range shifted upward.

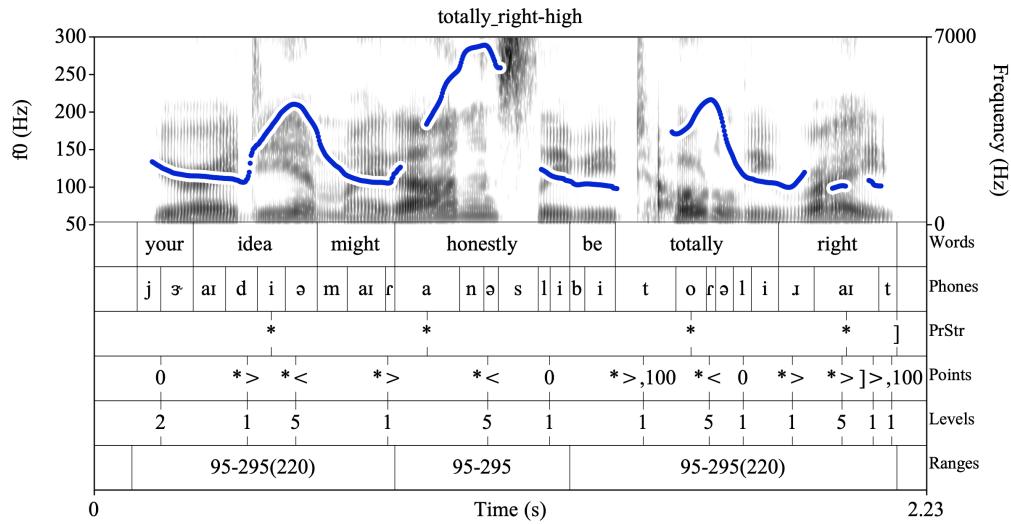
Parentheticals are often signalled via a different pitch range – commonly one that is lowered. The example below shows such a case, where the parenthetical phrase (“*who never listens*”), is produced in a pitch range (130-180Hz) that has both a lower ceiling than the preceding and following phrases, and also a more compressed range (only a 50Hz span compared to the neighboring ranges, which each span over 100Hz.) Here, the range domains correspond to the phrasing, as indicated in the PrStr tier.



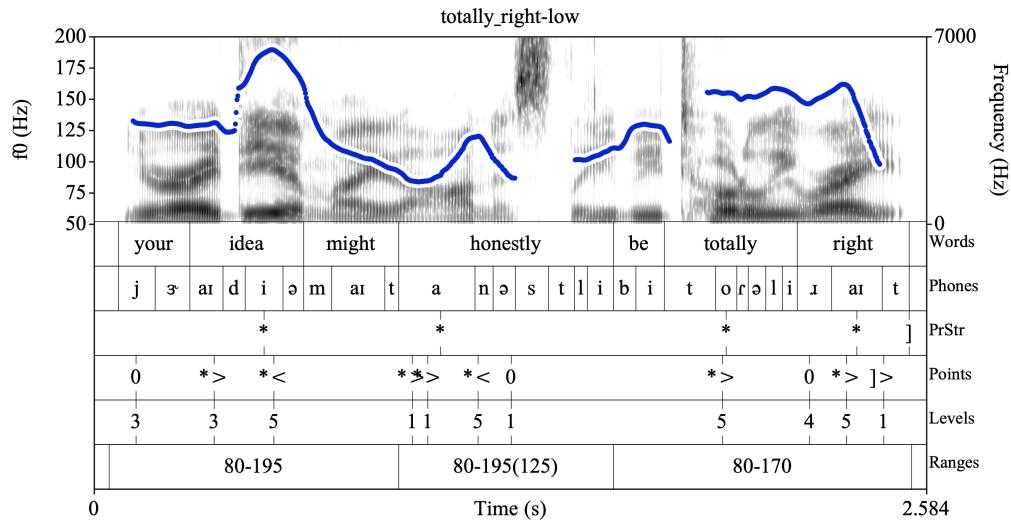
**Figure 1.49:** An example of a compressed parenthetical.

The following pair of examples illustrate that range changes signalling a parenthetical need not necessarily be accompanied by phrase boundaries in the PrStr tier. In the first example, the word *honestly* is produced in an expanded range compared to the preceding and following speech. And in the second text, the converse happens, with *honestly* produced in a compressed

range compared to the preceding and following parts of the utterance. In these examples, there are similar ranges on either side of the parenthetical word. Labelling the ranges for this type of example may be useful for researchers investigating the role of pitch in perceptual grouping, including non-local and global pitch relations (see, e.g., Ladd 1988, Geluykens & Swerts 1993, 1994, Brugos 2009, and Kentner & Féry 2013).



**Figure 1.50:** An example of a locally expanded range.



**Figure 1.51:** An example of a locally compressed range.

#### 1.4.8 How ToBI and PoLaR treat Ranges

A knotty issue in intonation labelling is how to deal with changes in a speaker's local 'operational range', both within and across utterances. This issue is critical, because it has profound implications for what counts as a High vs a Low intonational target. If you are familiar with

the ToBI system, you will recognize that ToBI addresses the question of local pitch range (and changes in that range) in several ways that do not involve explicitly labelling range information. We review some of these mechanisms in the following paragraphs, and compare them with PoLaR's approach to labelling changes in local pitch ranges (e.g., compression). In a following section we turn to PoLaR's approach to determining the intonational target scaled pitch level within a speaker's current 'operational pitch range'. (We are comparing and contrasting PoLaR with the ToBI annotation framework, given ToBI's widespread familiarity among intonationists. Note that several of the points made with respect to ToBI can be extended to other categorical/phonological annotation systems. Readers who are not interested in the relationship of ToBI to the PoLaR system may prefer to skip the present section.)

Since its inception, the phonologically-based ToBI system for annotating intonational prosody (cf. Beckman et al. 2005) has not been concerned with range changes, and intentionally has not provided tools for annotating ranges. This decision was motivated by a theoretical stance: the same phonological contour can be realized in different ranges, and the choice of range is assumed not to be a phonological distinction. However, in the years since ToBI was developed, extensive experience with intonational labelling has suggested that range changes may play a different and perhaps more important role than was originally assumed. As a result, PoLaR includes conventions for explicitly labelling the speaker's local f0 range, as it changes both within and between utterances. Here we discuss the phenomena of downstepping (compression where the pitch ceiling drops), upstepping (expansion where the ceiling rises), range shift (where both floor and ceiling move), and range ambiguity, which are handled differently in ToBI and PoLaR annotation, due to the differences in the goals of the two systems.

As a reminder, ToBI and PoLaR are mutually supplementary annotation systems, with different goals, which can be used in concert. PoLaR is, in a sense, a response to and an embodiment of insights acquired by many ToBI labellers over three decades of experience, which has (not surprisingly) identified a number of challenges in intonational labelling. We envision that the two approaches will be able to inform each other not only about the range of possible phonetic realizations of a given phonological specification of a tonal contour, but also about the possibility of identifying contrastive tonal phenomena not available for annotation in the ToBI system. PoLaR 1.0 represents our current proposal for the form that such a complementary acoustic-phonetic labelling system might take; feedback from users is actively sought and will be very welcome.

#### **1.4.8.1 How ToBI and PoLaR handle downstepping**

In our experience, the task of determining the local Range that a speaker is currently using, within his or her overall f0 range, is generally a reasonable one, although specific instances can be challenging. One such challenge arises because, in a sequence of High tonal targets, later High targets are frequently realized with a lower f0 than a preceding High target. There are two commonly distinguished sources for this phenomenon. The first is phonetic declination: the gradual lowering of the top of the speaker's f0 range (i.e., the f0 ceiling) over the course of speech. The second is perceived as contrastive (and thus arguably phonological), because the range drops steeply and suddenly. Thus intonational annotation systems are faced with a question of which examples (if any) of this type of event to annotate and (if so) how.

MAE\_ToBI addresses this issue by making use of a diacritic only in situations of downstep (using the ! diacritic before the relevant H symbol), and does not mark downdrift at all. Moreover,

the usage of the downstep diacritic is treated by ToBI as grammatically constrained: it can only be used when the previous High tone within the same intonational constituent of the relevant sort. In the current AM model (cf. Beckman et al. 2005:17–18, and references therein), the domain of downstep is the intermediate phrase, and the downstep label indicates that the pitch ceiling is lowered for the remainder of that intermediate phrase, specifying a narrower range. This grammatical analysis is beneficial in that it makes testable predictions, however it is not ideal for ease of labelling: it requires that labellers be trained to keep enough of the intonational model in working memory so as to be able to identify downstep only in such contexts while labelling. Moreover, it only allows labellers to keep track of certain types of range changes: only those that adhere to the grammatical definition of downstep.

In contrast, PoLaR Ranges are not subject to such constraints: in the PoLaR system, ranges can be annotated independently of the presumed phonological constituent structure of the utterance. As a result, what MAE\_ToBI regards as downstep (restricted to the ip domain) can in PoLaR be labelled in stretches of speech of any length or constituent structure, and even down-drift can be annotated with a different Ranges interval, if the labeller finds it to be a significant enough change over time. From a theoretical/analytical standpoint, this is useful for capturing cases where what appears to be downstep occurs across a phrase boundary, i.e. cross-phrase pitch range relations (cf. Ladd 1988, Brugos 2015), which are not possible to annotate while adhering to the grammar of MAE\_ToBI. From a practical standpoint, this is also beneficial in that labellers need not reference a grammar to decide when to track pitch compression by adding a new Ranges interval.

#### 1.4.8.2 How ToBI and PoLaR handle upstepping

A less common but still frequently-occurring phenomenon in American English intonation occurs when a High prominence is followed by another High with a higher f0 peak. This phenomenon can be thought of as upstepping. In MAE\_ToBI, at least, there is not a standard way to capture such a range relation. That is, a sequence labelled H\* H\* can be used in three different situations: when two successive High prominences are realized with approximately equal pitch height, when the second High is realized with a lower pitch that can be attributed to declination (as discussed above), or when the second High is noticeably higher. These possibilities cannot be distinguished in MAE\_ToBI, which is appropriate if they do not signal a phonological contrast. Nevertheless, it might be of considerable interest to know where they occur. ToBI systems developed for some other languages have implemented ways of capturing upstep; see, for example the discussion in Armstrong & Cruz (2014) of the use of the ; diacritic (inverted exclamation point) for annotation of varieties of Spanish and Portuguese. PoLaR takes a different approach to upstep, just as for downstep: such cases can be handled by annotating the beginning of a new and wider Range, to permit a distinction between upstepped and non-upstepped sequences.

#### 1.4.8.3 Range-labelling challenges addressed by PoLaR

Because range changes and phrase boundaries are tightly linked in the ToBI framework, some types of attested data are challenging to label. These problems include:

- 1) Range changes force annotation of a phrase break: When there is a big jump in range, either up or down, ToBI requires a phrase break, even if there is no other sense of disjunction.<sup>9</sup>

<sup>9</sup> An exception is made for cases where a disfluency is followed by the beginning of a new phrase before the disfluent

Because PoLaR allows the labeller indicate a range change within a phrase (or even within a word, as in Figure 1.46), it provides a tool for capturing within-phrase changes in f0 range when they occur.

2) Range changes prohibit annotation of a phrase break: In some cases there is a conflict between downstepping evidence that requires two targets to be in the same phrase, and other acoustic evidence that requires them to be in different phrases, because it suggests the presence of a boundary between them. ToBI labellers sometimes address this conflict by using the Break Indices 3- or 2, indicating conflicting evidence for the boundary, but in general the ToBI prohibition against downstepping across ip's means that if you want to keep track of perceived downstepping, you may need to violate your sense of phrasing.

Because PoLaR is agnostic about the theory-based constraint against downstepping across phrases, range change annotation is de-linked from the phrase/prominence annotation. Phrasal structure and prominence location are labelled in the PrStr tier, while Range is annotated in its own tier, so that a new Range can begin anywhere in an utterance.

3) Ambiguity between H and L targets: There are circumstances in which it is clear that a tonal target for a boundary or a prominence has occurred, but it difficult to determine whether that target is a H or a L tone. ToBI labellers are familiar with the uncertainty/ambiguity that arises when a pitch accent in a very compressed low range, often at the end of a phrase, might be a Low (L\*) or a downstepped High (!H\*). Similarly, sometimes an L\*+H label is required for a pitch accent even though the L\* accented syllable is realized with a higher f0 than the preceding syllable. PoLaR does not explicitly distinguish between H and L target tones, but rather uses the Range annotation and f0 values to determine which of 5 levels each target is realized with. As a result, it does not require labellers to make a phonological decision about H vs L, but instead provides the option of noting the compressed (or expanded) pitch range.

In sum, PoLaR provides a tool to address some of the challenges and ambiguities associated with labelling f0 range change phenomena that are always easy to deal with in more-phonologically-based approaches like ToBI. To the extent that these range changes are systematically controlled by speakers and attended to by listeners for communicative purposes, PoLaR's capacity to capture them may provide a useful supplement to phonological labels, as a means to systematically keep track of where pitch range is manipulated.

## 1.5 Levels (Advanced)

In PoLaR's present instantiation, the Levels are automatically determined by an algorithm, based on the hand-annotations in the Ranges and Points tiers. This algorithm divides the speaker's local pitch range (annotated on the Ranges tier<sup>10</sup>) into 5 equal quintiles, and each labelled point (annotated on the Points tier) is located in one of the 5 quintiles of the local Range.

In this way, the Levels tier provides labels for which the pitch heights of intonational targets across different f0 Ranges are normalized. The Ranges and the Levels labels capture information about where in the speaker's current f0 Range a tonal target is realized, reflecting the intuition that a tonal target with a Levels label of 5 in a lower or compressed range is conceptually just as high as a tonal target with a Levels label of 5 in a higher or expanded range. As

---

<sup>10</sup>phrase was completed, in which case the annotation %r (for 'reset') is used

<sup>10</sup>Note: the script that assigns Levels labels can interpret Advanced Ranges labels. In particular, range minima/maxima in parentheses (where they exist) override the values outside parentheses.

a result, the Levels tier and Ranges tier are intimately related. We turn now to a discussion of this inherent relationship, and how it interacts with some of the Advanced Ranges discussions addressed in the previous section.

### 1.5.1 The Influence of Levels and Ranges on One Another

In our experience, the task of determining the local Range that a speaker is currently using, within his or her overall f0 range, is generally a reasonable one, although specific instances can be challenging. Where there is doubt, a labeller can use the output of the Levels labelling algorithm to reevaluate their Ranges labels. Abstractly speaking, once the Points and Ranges for an utterance have been labelled and the Levels labeller algorithm has been run, a labeller can consider whether any two Points that have been assigned the same levels value are indeed in the same quintile band, or if they sound as if they reflect different tonal heights. If so, the labeller may want to consider that there is an additional Range that should be labelled, to accommodate the extra distinction. We will explore this more concretely below with three case studies.

**Advanced Levels Example 1: Levels-defined contours** As a first concrete example, consider again the examples in Fig. 1.32 and 1.34 from §1.4.2, repeated below.

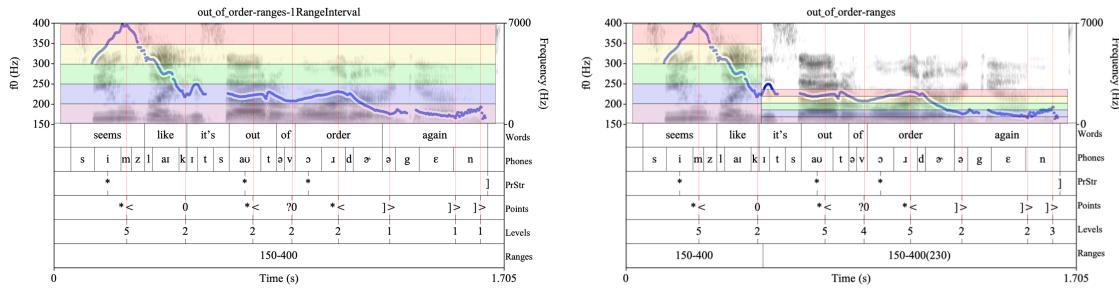


Figure 1.52: Different Range labels and their impacts on Levels labels.

Here the labeller may have begun with the Ranges label on the left, and then upon reconsideration decided that they heard the tonal height of “*seems*” as equally high in its local range as the tonal height in “*order*”. This is not reflected in the Levels labels on the left (“*seems*” has a level-5 peak while “*order*” has a level-2 peak.) This would have led to a revised Ranges label, as on the right, with two labelled intervals for different local pitch ranges (so that both have a level-5 peak in their stressed syllable). Whether the labels on the left or on the right are more appropriate depends on the labellers intuitions about the intonational contour: is it one with a high peak followed by many low pitch turning points (Levels: 5-2-2-2-2-1-1-1) or a high peak followed by a fall and then other high pitch turning points (Levels: 5-2-5-4-5-2-2-3).

Putting this more abstractly, if the pitch at two Points labels sound as if they are of the same pitch value, but they have different Levels values (given the Ranges labels), then a labeller ought to consider adding a new Ranges interval with different min/max values. (Also, if two Points sound as if they are of different pitch values but they have the same Levels value, a labeller ought to reconsider the number of Ranges intervals and their values.)

**Advanced Levels Example 2: Downstep or Downdrift** The issue of distinguishing among local Ranges within an utterance or across utterances can often be challenging, but it is made especially challenging in the context of pitch declination (due to downstep and/or downdrift). For example, when a pitch accent follows a sequence of pitch range compressions (i.e., after a string of ToBI !H\*s), it may be realized with a rather low f0 but it is not always clear whether there has been another application of pitch compression and it is high in its range (i.e., it is another ToBI !H\*) or whether it is low in its range (i.e., it is a ToBI L\*). While PoLaR does not make this intuition easier to access, PoLaR encourages the labeller to evaluate this question consciously, through the Levels and Ranges labels, and does not provide a simple binary choice of H or L. As an example, consider the different Levels labels that emerge based on the different Range labels in the figures below.

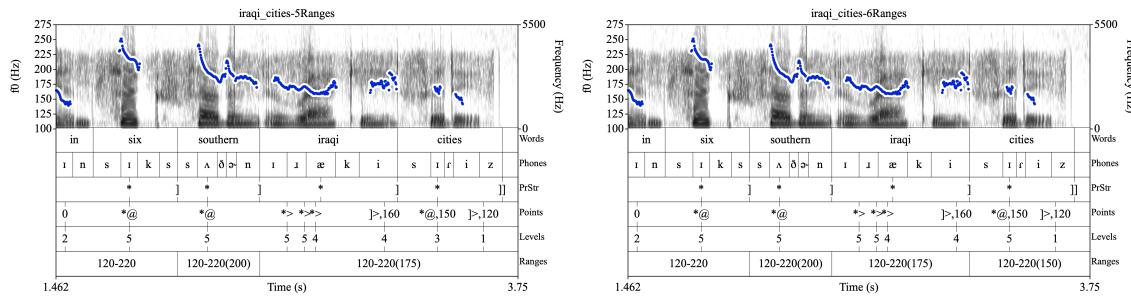
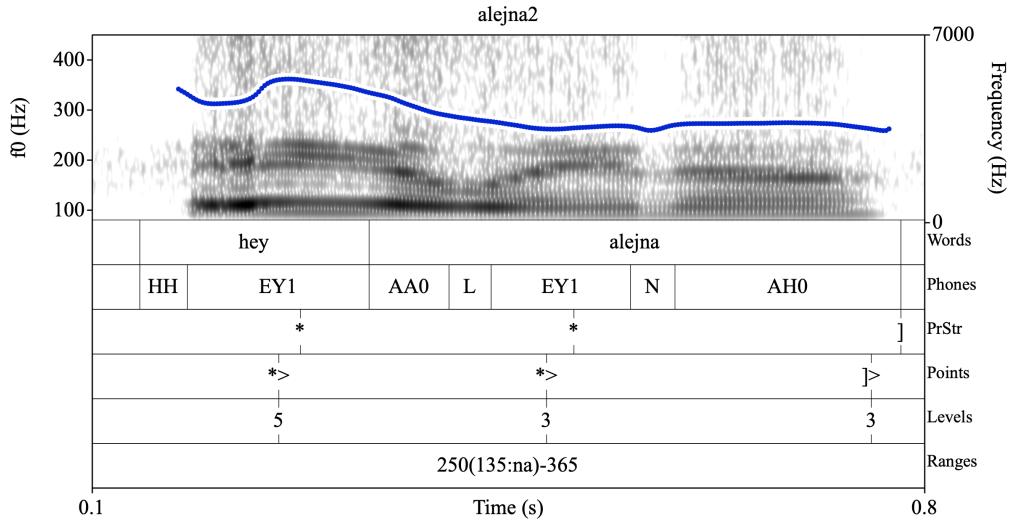


Figure 1.53: Interplay of Ranges labels and Levels labels.

If there is no additional range compression in the “*cities*” portion of the utterance (i.e., there is a single Ranges interval for “*iraqi cities*”), then the Levels label for the tonal target in the accented vowel comes out as a 3. If the labeller includes a final range compression for “*cities*”, then the same Levels label comes out as a 5. Ultimately, the labeller can use these different Levels values to inform whether they think the more appropriate Ranges labels (given their intuition about pitch height) includes this final pitch range compression or not. A labeller may also use their knowledge of intonational phonetics/phonology to inform which labels they prefer. Particularly, a labeller who believes the final pitch accent is simply a downdrifted phonological High might prefer the labels on the left, whereas a labeller who believes the final pitch accent is a downstepped phonological High might prefer the labels on the right. A different labeller might not analyze downstep or downdrift here at all, and might prefer Ranges labels where the Levels label in the accented vowel of “*cities*” comes out as a 1 or a 2 (as in, e.g., labels where there is just one Range of 120-220 for this entire portion of the recording).

**Advanced Levels Example 3: unattested pitch floor** A third example concerns cases where a speaker purposefully does not reach their pitch floor (or pitch ceiling, as may be in other parallel cases). An example of this is given in Figure 1.54, repeated from section 1.4.



**Figure 1.54:** Levels values where the attested f0 minimum differs from the intended pitch minimum.

A labeller may hear this recording as consisting of a high peak on “hey” falling to a mid or mid-high plateau for the remainder of the utterance. As such, the attested f0 minimum is around 250Hz, but this f0 minimum does not reflect low pitch. For this reason, the labeller can adjust the pitch minimum with an ‘:na’ label as done in this figure (cf. section 1.4.4). When it comes to identifying an appropriate value in the parentheses (here: ‘135:na’), a labeller can put a guess and run the Levels labeller script, iteratively, adjusting the parentheses value and running the Levels labeller, until the Levels value reflect the labeller’s intuition (here: that it is a mid-range pitch, i.e., a Levels value of 3).

In summary of this section about the interplay of Levels and Ranges, the point made here is that Levels (plus a labeller’s intuition and intonational model) may influence the number of Ranges that ought to be annotated. PoLaR’s flexibility (in that it does not limit its labellers to any particular grammar of intonation) means that there is no guidance on which labels are appropriate, but instead labellers are encouraged to actively check whether their PoLaR labels match their perception and beliefs.

### 1.5.2 Levels and Phonological Categories

In contrast to Autosegmental-Metrical phonology (as embodied in, e.g., the ToBI annotation system), which has as a core claim that tonal values are only either High or Low, PoLaR’s Levels labels differ in allowing a greater number of values. At the same time, PoLaR is not meant to indicate that there are more than two phonological categories of pitch height. Instead, one could (as the present authors do) adopt the premises (a) that the contrastive phonological categories are H vs L, and (b) that this contrast can be realized in various ways (on a five-point scale), across pitch ranges. As such, PoLaR labelling can serve to enable testing of hypotheses related to how H and L phonological categories are mapped onto particular phonetic targets or f0 Levels in the current f0 Range – with perhaps different mappings in different prosodic and/or pragmatic/semantic contexts.

One way in particular that this could be explored is through additional scripts that automatically map Levels values onto “H” and “L” categorical labels, using different algorithms to test which aligns best with human pitch perception. For example, a script might map Levels values 1 and 2 onto “L”, Levels values 4 and 5 onto “H”, with “3” mapping onto “H” when it’s above the Range interval’s midpoint and “L” when it is below. This particular mapping algorithm is not intended as the stance PoLaR takes, but rather represents a way that these more gradient Levels labels could be used for reconstructing labels from AM traditions. This reflects a broader idea that PoLaR may be able to facilitate the generation of categorical, theory-specific labels, on the basis of conceptually simpler and script-generated labels. (The PoLaR labelling team aims to further develop the PoLaR supplementary tools —currently found as a Praat plugin on polarlabels.com— to include a script to generate first-pass attempts of MAE\_ToBI labels from PoLaR labels for American English recordings.)

### 1.5.3 Outlook for Levels Annotation

In sum, the PoLaR annotation system makes explicit the claim that speakers can dynamically adjust their pitch range (even within a prosodic phrase) in such a way that what constitutes a “high” or “low” pitch can vary over the course of an utterance. A PoLaR labeller ought to check whether their Ranges tier is properly labelled by checking whether the Levels labels produced accord with intuitions and analysis of what counts as categorically High or Low.

At this point in the development of PoLaR, there are a number of issues related to Levels within the speaker’s current Range which have not been resolved and are under active discussion. We name three below.

*How should the Range be divided up algorithmically into levels?* Currently, as noted above, the local Range is divided into five equal quintiles. We anticipate that this algorithm may need adjustment, as more discoveries are made. For example, this number of divisions might be either too small or too large, or it might be inappropriate to treat these levels as equally-sized, in terms of pitch space. There are many plausible alternatives to this approach. For example, one alternative might be to vary the width of the quintiles so that the middle section of the current Range, designated as ‘3’, is wider than ‘1’ and ‘5’, or vice versa. Another possible algorithm might divide the pitch space into ten levels. Because the Levels-labelling algorithm is easy to adjust (via editing the PoLaR plugin scripts), such adjustments should be made as more empirical findings emerge (perhaps on the basis of studies that use Levels labels) about the most appropriate approach.

*Which units are best for specifying f0 ranges and levels?* As currently implemented, Ranges are expressed in Hertz (Hz), and Levels labels are also calculated on the basis of Hz. However, we recognize that there may be advantages to using logarithm-based semitones (e.g., differences among speakers are lessened; the human auditory system has been described as semitone-sensitive [Nolan 2003 and references therein]), or other methods such as z-score-based normalization. Just as for questions about how to divide up the pitch space, if changes in the units used for the analysis algorithm are desired, they are easy to implement in the PoLaR framework (and in the PoLaR plugin scripts), and changes can be made as empirical findings emerge about the most appropriate approach.

*What is the relationship between H vs L and 5 different Levels in a Range?* The fact that 5 different levels within a range appear to be appropriate for capturing significant perceptual differences (at least informally) raises the possibility that different Levels can carry functional loads even

though they are not phonologically contrastive. The PoLaR labelling system has the potential to enable closer examination of this question.



# Bibliography

- Ahn, Byron & Veilleux, Nanette & Shattuck-Hufnagel, Stefanie. 2019. Annotating prosody with PoLaR: Conventions for a decompositional annotation system. *The Proceedings of 19th International Congress of Phonetic Sciences*.
- Armstrong, Meghan & Cruz, Marisa. 2014. The intonational phonology of Iberian Spanish and European Portuguese. In Amaral, Patrícia & Carvalho, Ana Maria (eds.), *Portuguese-Spanish interfaces: Diachrony, synchrony, and contact*, 151–174. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Beckman, Mary E. & Ayers, Gayle M. 1997. Guidelines for ToBI labelling. *The OSU Research Foundation* 3. Available at [http://www.ling.ohio-state.edu/~tobi/ame\\_tobi/labelling\\_guide\\_v3.pdf](http://www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf).
- Beckman, Mary E. & Hirschberg, Julia & Shattuck-Hufnagel, Stefanie. 2005. The original ToBI system and the evolution of the ToBI framework. In Jun, Sun-Ah (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 9–54. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199249633.003.0002
- Beckman, Mary E. & Pierrehumbert, Janet. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* 3. 255–309.
- Brugos, Alejna. 2009. Non-local pitch range relationships in read and elicited speech. Presented at Workshop on Prosody and Meaning, Barcelona, Spain.
- Brugos, Alejna. 2015. *The interaction of pitch and timing in the perception of prosodic grouping*: Boston University dissertation. doi:<https://hdl.handle.net/2144/14012>
- Brugos, Alejna & Langston, Alison & Shattuck-Hufnagel, Stefanie & Veilleux, Nanette. 2019. A cue-based approach to prosodic disfluency annotation. In Calhoun, Sasha & Escudero, Paola & Tabain, Marija & Warren, Paul (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, Canberra, Australia: Australasian Speech Science and Technology Association Inc.
- Cole, Jennifer & Mahrt, Timothy & Hualde, José I. 2014. Listening for sound, listening for meaning: Task effects on prosodic transcription. *Proceedings of Speech Prosody* 7 859–863.
- Dehé, Nicole & Wichmann, Anne. 2010. The multifunctionality of epistemic parentheticals in discourse: Prosodic cues to the semantic-pragmatic boundary. *Functions of Language* 17(1). 1–28.
- Dilley, Laura & Shattuck-Hufnagel, Stefanie & Ostendorf, Mari. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24(4). 423–444. doi: 10.1006/jpho.1996.0023

- Geluykens, Ronald & Swerts, Marc. 1993. Local and global prosodic cues to discourse organization in dialogues. In *ESCA workshop on prosody*, 108–111.
- Geluykens, Ronald & Swerts, Marc. 1994. Prosody as a marker of information flow in spoken discourse. *Language and speech* 37(1). 21–43.
- Itô, Junko & Mester, Armin. 2012. Recursive prosodic phrasing in Japanese. In Borowsky, Toni & Kawahara, Shigeto & Sugahara, Mariko & Shinya, Takahito (eds.), *Prosody matters: Essays in honor of Elisabeth Selkirk*, 280–303. London: Equinox Publishing.
- Kentner, Gerrit & Féry, Caroline. 2013. A new approach to prosodic grouping. *The Linguistic Review* 2013 30. 277–311.
- Kim, Jiseung. 2020. *Individual differences in the production and perception of prosodic boundaries in American English*: University of Michigan dissertation.
- Ladd, D. Robert. 1988. Declination “reset” and the hierarchical organization of utterances. *The Journal of the Acoustical Society of America* 84(2). 530. doi:10.1121/1.396830
- Ladd, Robert. 2008. *Intonational phonology*. Cambridge: Cambridge University Press 2nd edn.
- Liberman, Mark & Pierrehumbert, Janet. 1984. Intonational invariance under changes in pitch range and length. In Aronoff, Mark & Oehrle, Richard (eds.), *Language sound structure*, 157–233. MIT Press.
- Nespor, Marina & Vogel, Irene. 1986. *Prosodic phonology*. Dordrecht, Holland ; Riverton, N.J.: Foris.
- Nolan, Francis. 2003. Intonational equivalence: an experimental evaluation of pitch scales. In *The proceedings of 15th international congress of phonetic sciences (ICPhS)*, 771–774. International Phonetic Association.
- Selkirk, Elisabeth. 1981. On the nature of phonological representation. In Myers, Terry & Laver, John & Anderson, John (eds.), *The cognitive representation of speech*, 379–388. New York: North-Holland Publishing Company.
- Selkirk, Elisabeth. 2011. The syntax-phonology interface. In Goldsmith, John A. & Riggle, Jason & Yu, Alan (eds.), *The handbook of phonological theory*, Blackwell Publishing 2nd edn.
- Shattuck-Hufnagel, Stefanie & Cutler, Anne. 1999. The prosody of speech error corrections revisited. In Ohala, John J. & Hasegawa, Yoko & Ohala, Manjari & Granville, Daniel & Bailey, Ashlee C. (eds.), *The proceedings of 14th international congress of phonetic sciences (ICPhS)*, 1483–1486.