

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

Contents	i
List of Figures	ii
List of Tables	ii
1 Usage Cases for PoLaR	1
1.1 Capturing systematic variation	1
1.1.1 Exploring the Intonational Phonetics-Phonology Interface	1
1.1.2 Slopes in f0 movements	3
1.1.3 Intonational Variation across Speakers/Varieties/Contexts	6
1.1.4 Patterns Related to Pitch Ranges	9
1.1.5 Tonal Alignment and Tonal Center of Gravity	11
1.2 Issues in Labelling Minimized by PoLaR's Design	13
1.2.1 Disagreements and Uncertainty Minimized by PoLaR's Design	13
1.2.2 Unlabellable Cases Minimized by PoLaR's Design	16
Bibliography	21

List of Figures

1.1	The final low plateau and low plateau with an extra fall have conventionally been labelled as L-L% in MAE_ToBI.	2
1.2	A contour with two High pitch accents (MAE_ToBI: H*) can variably have a plateau between corresponding peaks or various degrees of sagging between them.	2
1.3	Two hypothetical utterances that differ primarily in pitch range; note different vertical scales.	4
1.4	Advanced PoLaR labels for a hypothetical utterance.	4
1.5	The comma override label on the final Points tier object allows the phrase-final movement to be analyzed as falling, potentially avoiding the need to exclude this recording from analysis.	5
1.6	Two recordings of the same line by two speakers.	7
1.7	Pseudo-categorical labels (3rd tier from the top) generated by a script in the Praat plugin, for <code>librivox_recording-1</code> and <code>librivox_recording-2</code>	8
1.8	The gold dot indicates perceived alignment of rise-fall contour when rise is scooped, while the blue dot indicates perceived alignment when rise is domed. These shape changes influence the perceived alignment even though the turning point remains constant.	12
1.9	MAE_ToBI labellers may disagree or be uncertain about whether what's most appropriate for the initial high pitch is a H* label, a %H label, or no label at all. The different possible labels are indicated on the three "ToBI Alt" tiers.	15
1.10	MAE_ToBI labellers may disagree or be uncertain about whether a label of L+H* or H* is more appropriate for " <i>indeed</i> ". The different possible labels are indicated on the two "ToBI Alt" tiers.	16
1.11	A steady rise to a peak in the final (unstressed) vowel of " <i>menu</i> ", which is not perceived as prominent or at a phrase boundary.	18
1.12	A steady high through the final vowel of " <i>number</i> ", which is not perceived as prominent or at a phrase boundary.	18

List of Tables

Chapter 1

Usage Cases for PoLaR

Having reviewed the PoLaR annotation framework, we turn now to some contexts where we see PoLaR as being particularly useful.

A practitioner of prosodic labelling often has a persistent sense that (i) there are systematic differences (possibly implementational, possibly contrastive) that are not being captured; (ii) difficult-to-label tokens may be pointing to ways in which the theoretical framework can be revised or extended; and (iii) aspects of inter-labeller disagreement could be (and ought to be) lessened. Moreover, the practitioner may have the sense that, generally, these concerns could be addressed if there were a) appropriate phonetic labels that more transparently reflect observable aspects of the signal, and b) more widespread consideration of the possibility that different speakers, listeners (and labellers) may use the different cues to prosodic constituent structure and prominence in different ways.

A wide variety of research contexts are to some extent entangled in all of these issues; in this section we briefly discuss some of the ways in which PoLaR can help to address them, largely because it provides more information about the phonetic realization of a prosodic contour. These advantages include capturing systematic variation (e.g. in pitch slopes, pitch contours, pitch ranges, and cross-speaker differences in implementing a phonological category), and a minimization of labelling-related issues (e.g., labeller uncertainty or inter-labeller disagreement). We also discuss the ways in which this annotation framework makes the development of labelling skills more accessible to new labellers, and how it can be used in contexts where categorical labels are not suitable (such as exploration of an understudied variety or language).

1.1 Capturing systematic variation

1.1.1 Exploring the Intonational Phonetics-Phonology Interface

PoLaR labelling can be thought of as capturing a set of acoustic/perceptual characteristics of intonation in a way that is different from phonological labelling systems (like MAE_ToBI) that aim to capture discrete phonological units. A researcher using the latter sort of labelling may benefit from the additional information captured by PoLaR labels in two ways. First, PoLaR labels capture more details about the relevant raw acoustic cues (e.g., going beyond “High” and “Low” to include scaled pitch level and the size of the pitch range it occurs in), which can be useful to help understand how phonemic categories are phonetically realized. Second, PoLaR can be used to systematically explore the predictions and/or adequacy of a particular categorical

labelling system, with respect to its adopted phonemic inventory or phonological model.

As a concrete example for this general description, consider two cases where the conventions of MAE_ToBI labelling (by design) collapse a range of acoustic realizations. In each of Fig. 1.1 and Fig. 1.2 below, multiple possible f_0 tracks are depicted, and the same set of MAE_ToBI labels (/H* H* L-L%/) would be used to capture them.

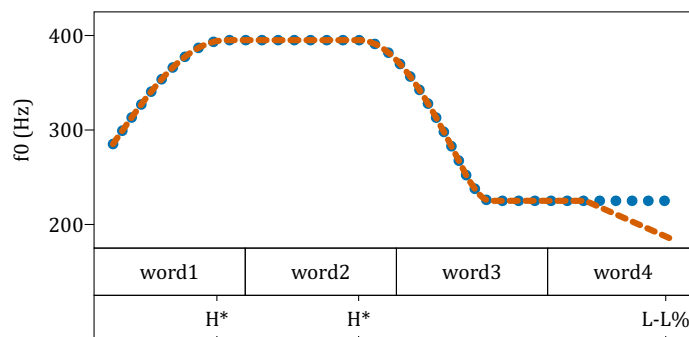


Figure 1.1: The final low plateau and low plateau with an extra fall have conventionally been labelled as L-L% in MAE_ToBI.

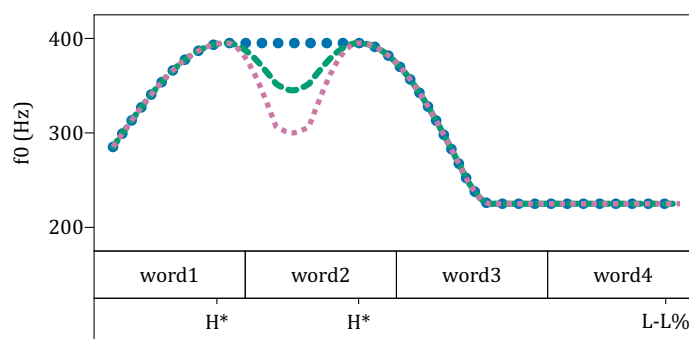


Figure 1.2: A contour with two High pitch accents (MAE_ToBI: H*) can variably have a plateau between corresponding peaks or various degrees of sagging between them.

Using PoLaR's (mostly phonetic) labels alongside phonemic labels like ToBI's could help explore systematicity in the phonetic variation within these phonemic labels. For example, in the case of Fig. 1.1, PoLaR Points, Levels, and Ranges labels would differ between the two contours, potentially uncovering systematicity in how L-L% is realized. In the case of Fig. 1.2, such sagging has been argued to be surface variation of the same phonological sequence of H*s; PoLaR Points and Levels Ranges labels may help to uncover whether there is systematicity about the degree, timing, and depth of such sags.

In addition, such work (or even work where PoLaR labels alone are annotated) may even possibly uncover new phonemic distinctions. For example, there may be phenomena that have been treated as category-internal variation (e.g., L-L% realized as either a final fall or low plateau, or various degrees of sags between H*s) where the variation turns out to be contextually-determined; and if the relevant contextual variables are semantic/pragmatic, this could be analyzed as an undiagnosed phonemic distinction, which in turn could be used to argue for readjusting the intonational phonemic inventory of the language.

More broadly, using PoLaR (alongside independently annotated categorical labels) facilitates

exploring questions about what sort of variation there is within-category, in the domain of intonation. In addition to the sorts of questions asked for the figures above, side-by-side usage of PoLaR and phonemic labels could explore questions like:

- At what rate do accents like L* or H* get realized with each of the possible Levels values?
- How do peaks/valleys (as labelled on the Points tier) align with Phone segments that are marked with a pitch accent label like H* or L*?
- How are the beginnings of phrase-final boundary plateaus (ToBI H- and L-) timed with respect to Phones segments and/or PrStr * and] labels?
- etc.

In other words, PoLaR enables users to systematically annotate a variety of cues (including those that a labeller attends to when choosing each phonological label) in a way that can be used to track potentially systematic variation (even in post-hoc analysis). To sum up why this type of more detailed labelling is important: it is impossible to discover which differences are systematic if the differences are not tracked (e.g., with annotation).

1.1.2 Slopes in f0 movements

As discussed above, many types of variation are potentially systematic and context-governed but are not possible to capture with a phonological labelling system, which by definition asks its users to lump together a variety of realizations under each of a particular set of predefined categorical labels. A categorical labelling system would not, by design, distinguish all degrees of phrase-final rises or of phrase-final falls. However, it may still be an open question how many communicatively different degrees of rise or of fall exist. For example, does the difference between boundary-related low flat f0 and a low but falling final f0 (as depicted in Figure 1.1, above) convey a difference in meaning, or are these likely to be used in different contexts? If so, the categorical labels ought to reflect such a difference. However, answering questions like this, about the phonetics-phonology interface, requires keeping track of degrees of final rises/falls systematically.

The PoLaR Levels tier (and indirectly the Points and Ranges tiers) can facilitate just this. For example, these labels distinguish a rise from level 3 to level 5 and a rise from level 1 to level 5, or a low flat final f0 from one with an additional final fall on the last syllable. PoLaR makes these scaled levels of pitch explicit, which facilitates exploration of how many categories of rise or fall need to be described by the phonology, by making it possible to look at how the acoustically-different realizations cluster. Doing this kind of tracking may be familiar to many researchers, who measure phrase-final f0 movements as occurring over intervals of some particular operationally-defined size. Though this can of course be done without PoLaR, PoLaR can streamline this process. Namely, since Points/Levels tier labels already track where f0 changes slope, they are ideal candidates for defining the endpoints of the interval for phrase-final f0 movements. In summary, using PoLaR labels to define endpoints for f0 slope can add a layer of methodological consistency to defining the interval for such f0 movements.

In addition, given that Levels labels are “scaled” (i.e., normalized, based on the Ranges tier labels), the rate of change calculated from Levels may facilitate comparison across individuals and/or recordings. Imagine the two hypothetical utterances shown below:

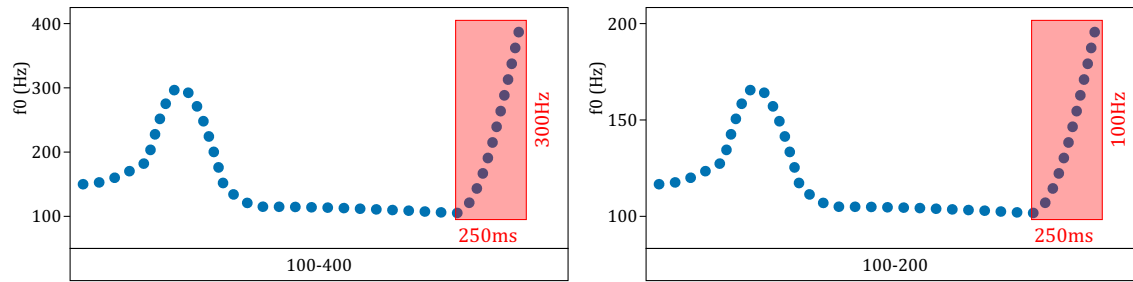


Figure 1.3: Two hypothetical utterances that differ primarily in pitch range; note different vertical scales.

The difference between the first and second hypothetical utterances is the f0 scaling: as shown by the scale at the left side of each panel and the range markers at the bottom, the former has a pitch range of 100-400Hz, while the latter has a pitch range of 100-200Hz. In terms of PoLaR, these two utterances would be annotated identically, save for Ranges tier. In both cases, the two phrase-final Levels labels would be 1 and 5 (a rise from low to high), Occurring .25sec apart, the slope would be 16 Levels/sec. However, if slope were calculated based on absolute f0 values, the two would have very different slopes: the former would have a slope of 1200Hz/sec, and the latter would have a slope of 400Hz/sec. In other words, because of how Levels labels are calculated relative to the pitch range, Levels labels inherently normalize rate of change in pitch. This is not to say that Levels/sec is a better measure of pitch slope —indeed, Hz/sec (or st/sec) may be more appropriate, at least in some cases— but rather that PoLaR Levels offer a straightforward way to compare pitch contours without additional data wrangling.¹

Beyond this, if Advanced PoLaR labels are used —in particular Advanced Points labels— then PoLaR provides another advantage: clearly keeping track of which f0 turning points are analyzed as related to prominences and which are related to phrase boundaries. Thus the f0 slopes between intonational events of different types can be tracked. For example, imagine we had the Advanced labels as in Figure 1.4:

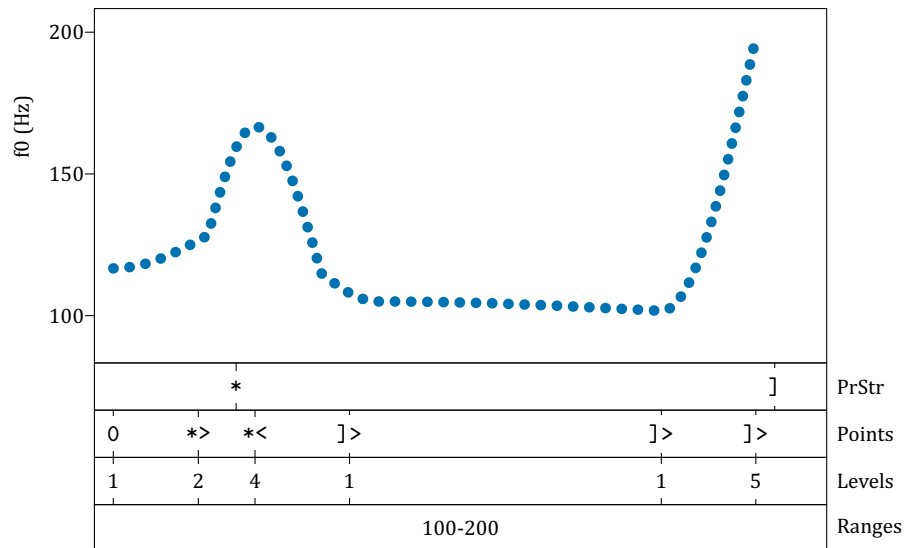


Figure 1.4: Advanced PoLaR labels for a hypothetical utterance.

¹Note that additional time normalization may be desirable for comparisons in some cases, but that is an issue that PoLaR labelling does not directly facilitate.

Using these labels, we could measure slopes like the steepness of the rise between the prominence-related Points (the ones labelled here as *> and *<), or the steepness of the fall from the last prominence-related Point and the first phrase-related Point (the ones labelled here as *< and]>). As before, these are measures that may be familiar to intonationists, but PoLaR labels facilitate finding the relevant endpoints for the measures of slope.

An additional way in which PoLaR can be helpful for calculating f0 slopes is when it comes to files with poor f0 tracking. It is common for those working on f0 measures to need to exclude some number of recordings because of software errors, recording conditions, or microprosodic effects of segments or voice quality. However, because PoLaR labellers can supply comma-override labels in these contexts, many of these files with f0 tracking errors could still be used. For example, the Figure below exhibits f0 tracking errors during “out there”, during which the pitch is perceived to be falling.

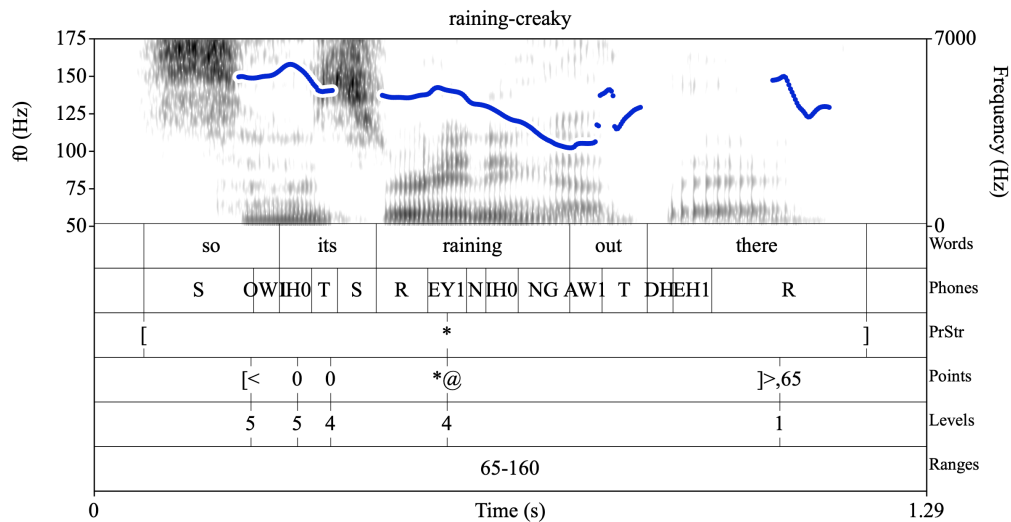


Figure 1.5: The comma override label on the final Points tier object allows the phrase-final movement to be analyzed as falling, potentially avoiding the need to exclude this recording from analysis.

Here the f0 at the Points label during “raining” is read as 140.8Hz at 0.547sec, with the final Points label is read as 65Hz (from the comma-override) at 1.290sec – a slope of approximately -102Hz/sec (or in terms of Levels, approximately -4 Levels/sec). Thus comma overrides can be used as a way to include recordings that might otherwise need to be excluded.

For analyses related to slopes of f0 movements, users are encouraged to employ the “Extract info from PoLaR tiers” functionality of the PoLaR plugin for Praat. It can run on an entire directory (or a single file), and for each object on the Points and Levels tiers, it outputs time alignment, f0 value, label, and much more information that can be useful for analysis. It also will respect comma-override labels, and use those values instead of direct measures. Using this and other scripts is briefly discussed in Ch. ?REF?, as well as in the plugin documentation.

PoLaR labels can systematically encode listener intuitions relating to the acoustics (such as

f0 slope), thus facilitating more targeted acoustic measures. Effectively, PoLaR labels provide more and/or different information than phonological labels do, providing valuable data for the exploration of questions about intonational phonetics and phonology.

1.1.3 Intonational Variation across Speakers/Varieties/Contexts

Another issue at the phonetics-phonology interface of intonation concerns variation across speakers in the phonetic realization of particular phonological categories. In the fields of phonetics, phonology, and sociolinguistics, a very active area of research explores variation in the realization of particular consonant/vowel categories, as well as how the number of phonological categories for consonants and vowels can differ across individuals or varieties. Given that transcription (both narrow and broad) is the first step in doing such comparisons, achieving similar goals in the domain of intonation requires an adequate transcription system. We propose that PoLaR can be useful for the narrower aspects of transcription, while other grammar-oriented transcription systems, like a ToBI system, may be useful for the broader transcription.

As an example, consider the steep-rise pitch accent in American English (in MAE_ToBI's broad transcription: /L+H*/). Burdin et al. 2018 reports on differences in the acoustics of and the frequency of /L+H*/ in five varieties of English spoken in the U.S. (Jewish English, African American English, Appalachian English, Midland U.S. English, and Southern U.S. English), finding that there are differences in “peak contour height, slope, and peak offset”. Though this work can be done without PoLaR, as discussed in the previous section, PoLaR labelling can be helpful in this domain. Points and Levels tier annotations can be especially useful in calculating f0 height and slope associated with particular phonological events such as pitch accents – especially if Advanced Points labels are used. Points tier annotations can also be used for calculating peak offset, given appropriate Phones tier labels for segment boundaries. (For an example of work that uses PoLaR to explore acoustics of steep-rise pitch accents across contexts and individuals, see Holliday 2021.) In addition, if Advanced Points labels are used, one could also track variation in how many turning points are associated with a pitch accent – perhaps some rises involve only two points, but perhaps others (e.g., “scoopy” rises) require more points, or perhaps this varies across varieties. In other words, PoLaR allows researchers to investigate phonetic variation without piecing together annotation methods and ad-hoc practices for phonetic measurement, and without making assumptions that all English varieties share categories, which we return to momentarily.

In addition, Levels labels serve as a transcription of pitch heights beyond raw f0 values that is a narrower transcription than broad categories like H or L. This narrow transcription can be used as a way of clustering different productions together, similar to how narrow-transcription IPA symbols are used to keep track of allophones and their usage in studies on the phonetics-phonology interface or sociolinguistic variation. In this way, one could track the rates at which /L+H*/ transcriptions are realized as [1-5] rises, [2-4] rises, [3-5] rises, etc, and which contextual variables help to predict how big (in scaled terms) of a rise is used. To be clear, PoLaR can be used alongside phonological systems like ToBI to track which allophones exist and what sociolinguistic variables or contextual conditions from the phonology matter for describing the distribution of the different forms that a phonological object (e.g., a pitch accent) can take.

Below are two recordings of the same line by different speakers, serving as a brief exposition of how PoLaR can be helpful for research questions like this:

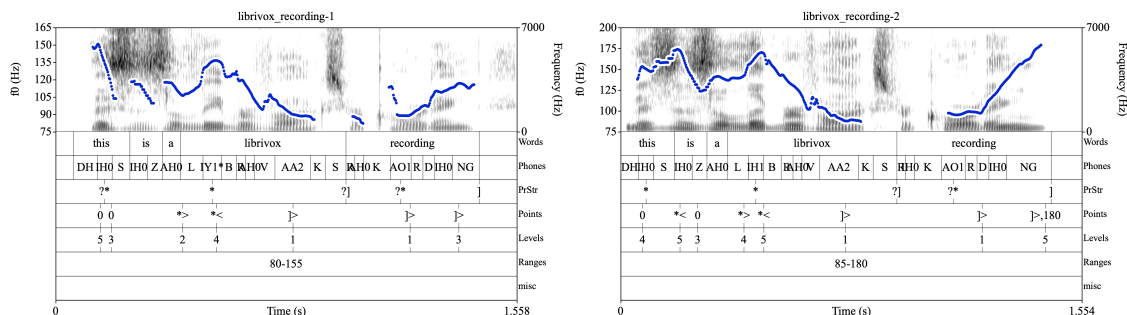


Figure 1.6: Two recordings of the same line by two speakers.

Here we see both speakers have Points tier labels associated with a * on “Librivox”. Immediately we see the two are different in that the first speaker goes up two Levels (2-4) while the second only goes up one (4-5). At the same time, the latter has a slightly bigger local pitch range (95Hz) as compared to the former (75Hz). Comparing measurements based on these labels, we can also see that the Levels-based slope of the rise in the first one is slightly steeper (17.4 Levels/sec) than the rise in the second (14.9 Levels/sec), and that the peak in the latter is aligned earlier in the first example (towards the end of the stressed vowel interval) than in the second (just after the stressed vowel interval). These sorts of measurements could be calculated easily repeated across a large number of recordings, by making use of the “Extract info” functionality of the PoLaR plugin for Praat (see previous section) and using spreadsheet formulas or scripts for statistical analysis.

Let us return now to the idea of exploring variation in /L+H*/ productions across different varieties. One issue that Burdin et al. faced (p.c.) is that the definitional boundaries of the L+H* category that they appealed to were crafted for mainstream US English; as such, it is potentially problematic to use this label for these other varieties, when varieties can have mergers or splits with respect to categories (cf. variation in American English with respect to *caught*~*cot* or *merry*~*Mary*~*marry*). In other words: is it a problem to assume there is an L+H* category that exists across varieties of English? And if so, how can we define the intonational form of that category so that it can be identified and compared across these varieties? To address this issue—which results from a top-down approach where a pre-established definition is used to pick out what should be acoustically compared—PoLaR labels can be used to define which intonational movements are similar enough to be compared. In other words, PoLaR labels (built from the bottom-up) can be used alongside some notions of what defines a category (e.g., a PrStr event and associated Levels values), in order to allow those elements that belong to the same category (according to those metrics) to be compared.

Generating “Pseudo-Categorical” Labels from PoLaR Labels

To be more concrete, we are advocating that “pseudo-categorical labels” (resembling A-M style categorical labels, like those used in systems like ToBI and IViE) can be created on the basis of PoLaR labels from the Phones, PrStr, Points, and Levels tiers, so long as the Points tier has Advanced labels. An example of such pseudo-categorical labels is given below (the third tier from the top, “Pseudo”, shows these labels):

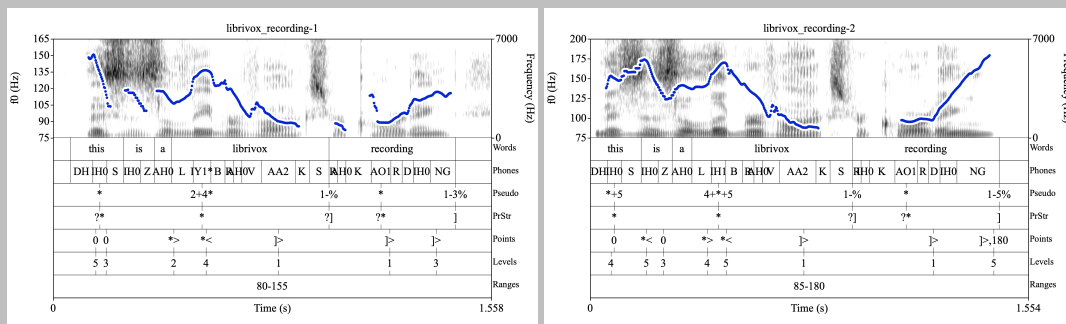


Figure 1.7: Pseudo-categorical labels (3rd tier from the top) generated by a script in the Praat plugin, for librivox_recording-1 and librivox_recording-2.

Pseudo-categorical labels such as these can be especially useful in keeping track of the distribution of allophones, as described earlier in this section. In the examples above, assuming each instance of “librivox” is labelled /L+H*/, we could say that /L+H*/ has allophones of [2+4*] and [4++5].

Of course using PoLaR labels to construct the boundaries of a category requires an algorithm. A preliminary attempt at implementing such an algorithm is coded into a script in the PoLaR plugin for Praat. The basics of this algorithm are as follows. We begin with prominence-associated labels. The script identifies the Phones-tier segments during which a * label occurs and any Points tier labels that are associated with that * (through <, >, or @ labels in the Advanced Points labels). For every *-associated Points label within the relevant Phones interval, the corresponding Levels label appears next to the * in the Pseudo tier label. For *-associated Points labels that precede the relevant Phones interval, the associated Levels label(s) is/are inserted, followed by a + (akin to ToBI complex tones). For example, during “librivox” in librivox_recording-1, there is a 2+4* label indicating that the Level 4 Point occurs during the *-marked segment, and it is preceded by a Level 2 Point that occurs before the *-marked segment. On the other hand, in librivox_recording-2, a 4++5 label occurs within “librivox”, indicating that neither Points label occurs within the *-marked Phone segment, and that a Level 4 Point occurs before the segment, and a Level 5 Point follows it.

As for boundary-associated labels, the script identifies the Phones-tier segments after which a] label occurs and any Points tier labels that are associated with that]. For every]-associated Points label within or after the relevant Phones interval, the corresponding Levels label appears before the % in the Pseudo tier label. For]-associated Points labels that precede the relevant Phones interval, the associated Levels label(s) is/are inserted, followed by a - (akin to ToBI phrase accents). For example, after “recording” in librivox_recording-1, there is a 1-3% label indicating that the Level 3 Point occurs during the]-marked segment, and it is preceded by a Level 1 Point that occurs before the]-marked segment. After “librivox” in librivox_recording-1 there is a 1-% label, indicating that there is a]-associated Level 1 Point before the phrase-final segment, and no]-associated Point during the phrase-final segment.

Of course this algorithm is merely one attempt at transforming PoLaR labels into (pseudo-)categorical labels, and the nature of the algorithm must be revisited, as more is learned about the phonetics-phonology interface for intonation.

Finally, individual speakers within a dialect may also show systematic differences in their phonetic intonational habits, just as they do for cues to segmental features. Given the increasing evidence that language users attend to and manipulate individual cues to phonological categories (and the values of those cues), it is likely to be fruitful to be able to capture these systematic patterns of variation for intonation. Because PoLaR is tightly tied to observable events and values in the signal, and goes beyond the labelling of phonological categories to include phonetic values, it provides a tool for annotating such patterns.

While the discussion in this section is oriented towards exploring interspeaker variation within a language, similar methodologies can be used to build up a tonemic inventory for a language for which there has been little to no work on the intonational phonology.

1.1.4 Patterns Related to Pitch Ranges

Though it is well known that pitch range is dynamic —both over the course of a single utterance and across utterances— less is known about the more precise ways in which they change. Once more precise annotation is kept, we can ask ifDo the different ways in which pitch ranges change correspond to different (linguistic) contexts. There has been some research on the topic of pitch range and its connection to a variety of linguistic disciplines — including prosodic phonology, syntax, semantics, pragmatics, sociolinguistics, and discourse structure— from empirical angles of both production and perception.

As discussed in § **?REF?**, PoLaR can be a useful tool for exploring the way that pitch ranges change. A core reason for this is that (as far as we know) no other intonational annotation systems beyond PoLaR requires explicit, systematic, and regular annotation of local pitch ranges and how they change. By including pitch range annotation as a core part of the labelling process, any PoLaR-labelled file can be used to investigate pitch range phenomena, even if that was not intended as a measurement for analysis by the original researchers who collected and/or labelled the data. Below we give a few examples of some research topics for which the Ranges labels could be used in analysis.

Below we give some examples of research that has yielded findings with respect to pitch ranges. After these paragraphs that provide an (abbreviated) overview of these topics, some abstract descriptions of ways in which PoLaR could be used to produce more nuanced results about pitch ranges will be provided. Finally, a sample recording is given with some more specific discussions about how PoLaR could be used to analyze the pitch range phenomenon.

In a variety of domains, there have been suggestions that changes in pitch ranges are tied to particular phenomena tied to linguistic meaning, broadly construed (e.g., semantically, pragmatically, and/or discourse structurally). This paragraph reviews one set of examples of pitch ranges ties to meaning, all from English. For example, yes/no questions have been found to have final rises that reach an extra high pitch, often higher than other highs in the utterance (cf. Pierrehumbert 1980), suggesting that there is pitch range expansion that includes (at least) the final boundary movements. (Though perhaps the expanded pitch range applies to the entire utterance, as might be suggested by the fact that initial f_0 is higher at the beginning of a YNQ than a declarative; Sicoli et al. 2015.) In addition, pitch ranges have been found to expand contexts expressing incredulity (Hirschberg & Ward 1992), surprisal/mirativity (Rett & Sturman To appear), and emphasis/focus (Xu & Xu 2005), or where the discourse topic shifts majorly (Hirschberg & Pierrehumbert 1986). Beyond these examples of expansion, pitch range compression —systematically lowered pitch accent f_0 peaks across phrases— has also been found

to be associated with parentheticals (Price et al. 1991, Dehé 2009) with discourse coherence/continuation (Beckman 1993, Bruce et al. 1997). More broadly, a variety of researchers have pointed to a role for pitch range relations in discourse segmentation, i.e. connecting phrases into coherent segments in discourses (Hirst 1993, Wichmann 2000, Hirschberg 2004, Lin & Fon 2011). Further research in these areas would benefit from PoLaR’s systematic labelling of pitch ranges. Similar connections between pitch range and meaning has been found in a variety of languages – a small sample of such findings are in the domains of focus (Mandarin: Xu 1999), parentheticals (French: Fagyal 2002), discourse coherence (Swedish: Hansson 2003, Carlson et al. 2005).

In addition to these domains that are tied to particular meanings, questions oriented to phonological structure and syntactic structure can be asked too. It has been suggested that changes in f_0 maxima for High pitch accents reflect hierarchical organization of prosodic phrases in multiple languages (Ladd 1988 for English, Féry & Truckenbrodt 2005 for German; van den Berg et al. 1992 for Dutch), and that such changes correspond to syntactic structure or cue grouping interpretation (Ladd 1988, 1992, Féry & Truckenbrodt 2005, Kentner & Féry 2013), even when timing cues may be ambiguous or conflicting (Brugos 2015). This direction of research relates to the question of the domain of downstepping, which has been said to be unable to apply across prosodic phrase boundaries in English (Beckman & Ayers 1997); however, Sturman 2019 suggests such cross-phrase downstepping is possible with at least some types of prosodic phrases (large ones; in MAE_ToBI terms: IP). Further research in this area is needed, and PoLaR is well suited to help, since Ranges are annotated completely separately from phrase boundaries.

As a final domain in which pitch range is explored, there are sociolinguistic and variationist findings tied to pitch ranges. For example, it has been suggested that African American English speech generally makes use of wider pitch ranges (marked by more falsetto and less downstepping; Wolfram & Thomas 2002, Thomas 2007), as compared to Mainstream US English. As another example, it has been found that newscasters and non-newscasters do not differ significantly in pitch ranges, but they do differ significantly in how much time is spent in different parts of their pitch ranges, and that this might be done as a way of encoding particular conversational goals that newscasters have (Gasser et al. 2019). PoLaR could again be useful to help keep track of not only Ranges (which could capture variation in usage of falsetto, downstep, and range size), but also Levels labels (which could capture how speakers make use of the same pitch range differently).

Let us turn now to the benefits of PoLaR for such explorations. Before proceeding, it must be mentioned that all the previously mentioned studies have achieved these results without PoLaR. While pitch range size can be tracked using directly observed f_0 minima and f_0 maxima, PoLaR provides a framework dealing consistently with common pitfalls related to this (cf. § ?REF? and § ?REF?). Further, the PoLaR framework provides tools for analyzing range annotations with respect to the annotations of prosodic structure (on the PrStr tier), pitch movements (on the Points tier), and scaled pitch values (on the Levels tier). Further, as described in (§ ?REF?), we can envision extensions to PoLaR to more directly associate Ranges tier labels to annotations of discourse or syntactic-pragmatic structure.

In many works on pitch range (including in references above), researchers have defined expansion and compression by comparing the f_0 height of pitch maxima across different utterances. However, since pitch ranges are defined by a ceiling and a floor, empirical questions remain

about the pitch range changes. While findings have been established in relation to changes in pitch ceilings, less is known about the extent to which pitch floors and range span might play a role. One possibility is that different types of pitch range changes (compression from one/both ends, expansion at one/both ends, or shifts up/down) are associated with different types of linguistic contexts or communities of speech.

As a case in point about the value of tracking both the floor and ceiling of pitch range, Dehé & Wichmann (2010) write that “the typical parenthetical prosody is often, although not always, a marked shift to a compressed pitch range”. The same PoLaR labels that could be used to explore lowered f0 ceilings could also be used to explore the following questions: when do parenthetical pitch ranges compress/expand, compared to preceding range? How similar/different to each other are the ranges on either side of the parenthetical? Does the range ceiling ever get lower than the preceding range floor? Does the range floor ever get higher than the preceding range ceiling? While some of these variables can indeed be measured without PoLaR, it would require researchers to establish standards and conventions of their own. On the other hand, PoLaR labelling already includes this on the Ranges tier, and provides labellers with systematic annotation guidelines. Moreover, PoLaR allows labellers to make use of their intuitions with respect to labels, in a way that other direct measurements might not.

Moreover, since it is known that pitch ranges are dynamic within an utterance, taking measurements of only pitch minima/maxima in an utterance restricts possible findings by preventing researchers from asking questions like how these pitch range changes are timed within an utterance (with respect to other prosodic events and/or particular words/morphemes). Labelling pitch ranges with PoLaR allows researchers to more directly address such questions via the greater detail about the magnitude and timing of range changes.

Finally, using PoLaR could create hypotheses based on observed patterns in pitch ranges, or to check hypotheses about patterns in pitch ranges, with Ranges (and other PoLaR) labels defining what is measured. For example: one might explore whether the quantitative values of Ranges (with the min and max as continuous variables) are tied to particular parts of the utterance (e.g., timing with respect to word or phoneme boundaries, or relation to PrStr events). As another example: those interested in exploring how pitch ranges are used (e.g., how a speaker uses pitch within a range) could explore that through the Levels values of turning points in the pitch track and/or by calculating the area under the curve in different utterances, with the f0 or Levels values defining the curve (i.e., the ceiling) and the Range min defining the baseline (i.e., the floor). In both of these cases, as well as in others, users are encouraged to use the Extract Info functionality of the PoLaR plugin for Praat to extract the relevant measures into a spreadsheet-style format for easy analysis.

1.1.5 Tonal Alignment and Tonal Center of Gravity

The alignment of the f0 contour over a prosodic event may cue various interesting aspects of spoken language. For example, the f0 alignment over the segmental string can be a cue to disambiguating the family of MAE_ToBI categories L+H* (late), H*, H+!H* (early) and L*+H (late) often implemented in a rise-fall-rise context. The alignment of a tonal event may also cue dialect distinctions or phonological allophones (that is, different f0 shape with same perceptual understanding). Various measures of alignment have been proposed and used (e.g. with respect to the nucleus onset or center or with respect to an f0 peak or other turning points). Interesting results from these investigations (e.g. Niebuhr et al. 2011, d’Imperio 2000) suggests

that shape in concert with Turning Point alignment must also be considered.

As an alternative to Turning Point based alignment, the Tonal Center of Gravity (cf. Barnes et al. 2010, 2012) is a global intonational measure that captures the perceptually salient aspects of alignment (TCoG-t) and scaling (TCoG-f). In the time domain, the TCoG-t is the f_0 -weighted point in time. In the frequency domain, TCoG-f is the average f_0 weighed by the intensity and recency with respect to f_0 movements. Together, the TCoG values are global measures that capture the impact of f_0 turning points, of f_0 contour shape and of the segmental/acoustic string on the perceived alignment of an f_0 contour. Thus the TCoG is a perceptually salient cue to f_0 alignment and scaling, arguably more so than any (potentially missing) directly-measured f_0 values such as peaks.

Figure 1.8 illustrates how the TCoG-t captures shape contributions in addition to the turning point locations. The TCoG (shown both in time and frequency here) varies even when the commonly-used rise-peak-fall turning points are held constant. As Barnes et al. (2010, 2012) have shown, these differences are perceptually salient, even when the precise point in time of the f_0 is fixed.

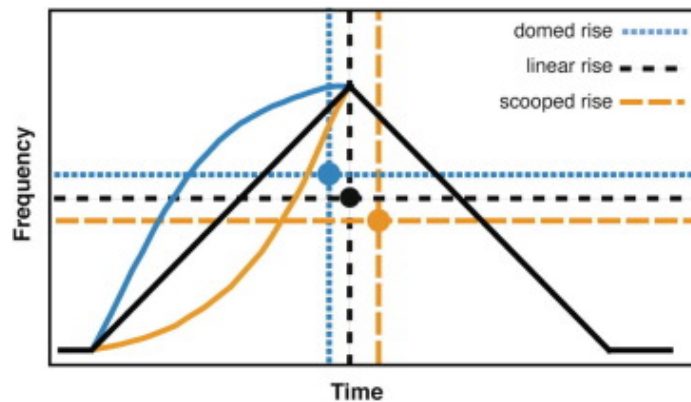


Figure 1.8: The gold dot indicates perceived alignment of rise-fall contour when rise is scooped, while the blue dot indicates perceived alignment when rise is domed. These shape changes influence the perceived alignment even though the turning point remains constant.

As an average over many f_0 values, the TCoG can furthermore be modeled to incorporate the impact of changes in f_0 salience over differences in intensity and/or sonority by including weighting factors. It has been suggested that F_0 perception is strengthened/more salient over areas of greater periodic energy (Albert et al. 2018), greater sonority (Barnes et al. 2014), and/or at the end of pitch movements (d'Alessandro & Mertens 1995).

However, once one concedes the need to account for both f_0 shape and alignment over the segmental string, there is still the open research question concerning the bounds of the salient interval. Specifically, over what interval is the weighted f_0 averaged? Two obvious candidates are (a) the segmental domain (related to the accented vowel/syllable) or (b) the intonational domain (related to deliberate prosodic event related f_0 movements) although it is possible that listeners integrate both.

In resynthesized laboratory stimuli adjusted to explore alignments, the segmental and f_0 turning point domains often are designed so that they overlap substantially. Furthermore, these stimuli do not take into account accommodations speakers might make to adjust to a segmental string (e.g., earlier or later f_0 movements around unvoiced regions or avoiding pitch

movements on non-accented syllables). However, using naturally occurring speech presents difficulties in calculating f0 values.

For example, (as mentioned extensively throughout this guide) human speech typically does not result in smooth and continuous f0 contours with clearly discernible turning points. In fact, turning points can be missing even during intervals thought to be the most important phonologically (e.g., peaks of High accents). For example, f0 points may be missing at critical locations due to a voiceless sound source, or the f0 track may take the form of a plateau. PoLaR provides two ways of capturing this missing information:

1. Annotators can use the “comma-override” in the Points tier to estimate the missing f0 values in crucial locations.
2. Annotators can place inflection points (i.e. in the dome or scoop elbow) in the Points tier to create distinct straight line approximations of the perceived f0, presumably adjusting the contour to align in the same way the TCoG does.

PoLaR labeling could be useful in providing data to investigate alignment and scaling behavior by allowing researchers to examine the TCoG in unconstrained (rather than laboratory) speech. Naturally occurring, unscripted speech carries variation in f0, the quality of the f0 contour, and the energy in segmental implementation which can provide variety to investigate accommodations that speakers make to preserve the meaningful prosodic cues. Using PoLaR labels would therefore enable automation of the TCoG calculations over large corpora, to explore the TCoG in utterances produced “in the wild”. This, in turn, could allow researchers to explore alignment and scaling in a larger variety of contexts.

1.2 Issues in Labelling Minimized by PoLaR’s Design

In addition to facilitating work on research questions about intonation in new ways, PoLaR’s design should minimize some common issues that arise in using existing intonational annotation systems (especially those that are categorical). In this section, we break down these issues into two fundamentally different types: those that are process-oriented (e.g., relating to disagreements and uncertainty) and those that are theory-oriented (e.g., relating to poorly understood and/or understudied intonational phenomena).

1.2.1 Disagreements and Uncertainty Minimized by PoLaR’s Design

As mentioned elsewhere in this monograph, PoLaR labels on different tiers are largely independent of one another, a design feature which has the potential to (i) facilitate labelling, (ii) significantly reduce labeller disagreements (e.g. by reducing both subjectivity and uncertainty), and (iii) guide discussion toward resolution of problems and new analyses. The independence of different tiers means that the labeller can deal with one tier at a time, and it is easier to learn how to label one tier than to master labels that bundle a variety of intonational characteristics together. Thus the independence of the tiers provides a low barrier to entry for novices. It also means that the overall labelling task can be divided among multiple labellers working in parallel on different tiers.

In addition to disagreements that may arise from bundled labelling, labelling discomfort sometimes arises because of uncertainty. To increase labeller confidence, PoLaR provides tools that can be used to check a candidate label. For example, the Straight Line Approximation tool pro-

vides a convenient mechanism for checking whether a particular Points label is required or not, and the Levels labeller tool can be used to check Ranges labels. Because PoLaR labels map relatively transparently onto signal characteristics, many aspects of labelling are quite mechanical (and for some aspects, fully automatic); this reduces uncertainty (and thus, disagreement). An exception to this is the categorical labels (i.e. PrStr labels), which are still quite abstract and require listener judgments. In the future, one way to avoid disagreements might be to source these labels from RPT (thereby replacing potential disagreements in categorical labels with crowdsourced labels serving as a kind of consensus label, and offering the additional benefit of capturing the degree of ambiguity perceived by a group of listeners).

Another advantage provided by the design of PoLaR is the separation between Basic labelling (which focuses on prosodically relevant acoustic characteristics and requires no or minimal interpretation) and Advanced labelling (which permits the labeller to add more details in a second pass, including more interpretive notations). The flexibility of the system also allows experienced labellers and researchers who are interested in particular aspects of the intonational signal to easily add relevant labelling tiers. These design features (separation and flexibility) have the potential to provide good agreement on the more transparent Basic labels, and to restrict disagreement to the Advanced labels, where the discussion may be more fruitful. We note that PoLaR focuses largely on cues rather than on the contrastive categories of intonational phonology; as a result, the fact that different speakers and listeners (and labellers) may interpret cues to phonological categories differently becomes less of a problem. The mapping between cues and phonological labels, which may be complex, is left for further analysis and discussion, while the annotation of the cues themselves may be more straightforward and less subject to disagreement.

To exemplify how PoLaR can reduce labeller disagreements and facilitate focused discussion, here we consider some examples of disagreements that often arise in alternative labelling systems that are designed for phonological labelling (e.g., MAE_ToBI), and consider how working with PoLaR might help. When such a case arises, the labels themselves direct labellers to specific aspects of the signal to discuss. For example, consider the utterance below, where labellers disagree about whether the high pitch during “it” ought to be labelled as H*, as %H, or not labelled at all.

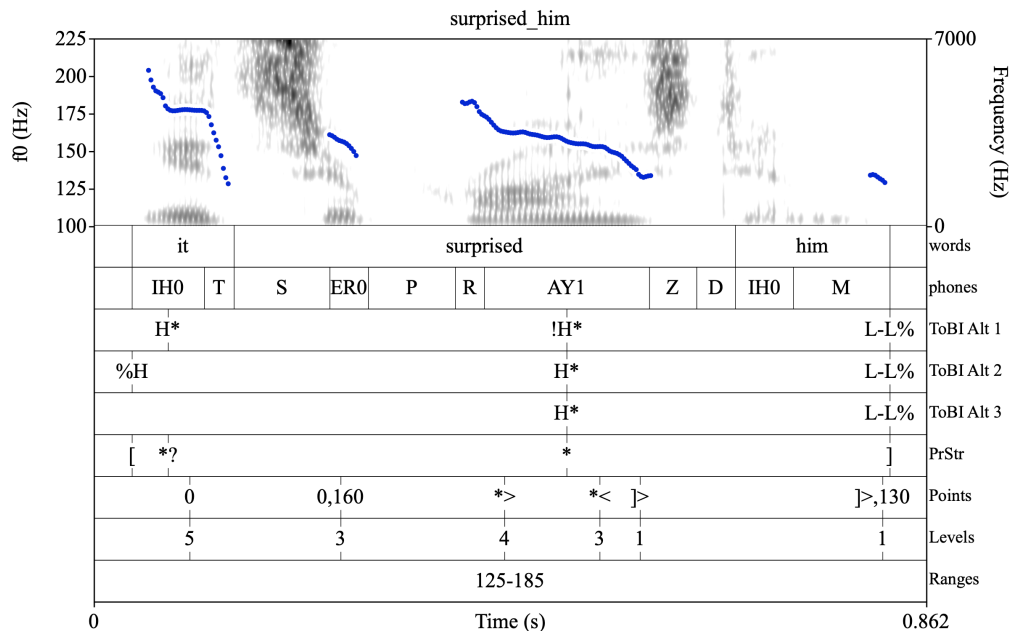


Figure 1.9: MAE_ToBI labellers may disagree or be uncertain about whether what’s most appropriate for the initial high pitch is a H* label, a %H label, or no label at all. The different possible labels are indicated on the three “ToBI Alt” tiers.

Is this because of different perceptions with respect to prominence (i.e., H* vs. %H or no label), or is it because labellers disagree whether the pitch is meaningfully at a high target (i.e., H* or %H vs. no label)? Instead of forcing annotators to choose an analysis before creating a complete set of labels, PoLaR allows the uncertainty of prominence to be captured as a *? on the PrStr tier, and the lack of commitment to a specific phonological analysis is captured as a \emptyset on the Points tier. For labellers who are more certain, the source of perceptual/analytical differences will be easier to identify and discuss, with PoLaR labels that pull apart the relevant issues. In this specific case, the perceptual issue of prominence will be clear through PrStr labels for “it” (i.e., * vs. *? vs. nothing), and the analytical issues can be captured as Advanced Points labels (i.e., <[vs. <* vs. \emptyset).

As a second example, labellers quite commonly disagree over—or are uncertain about—whether a particular label ought to be L+H* or H*, as below.

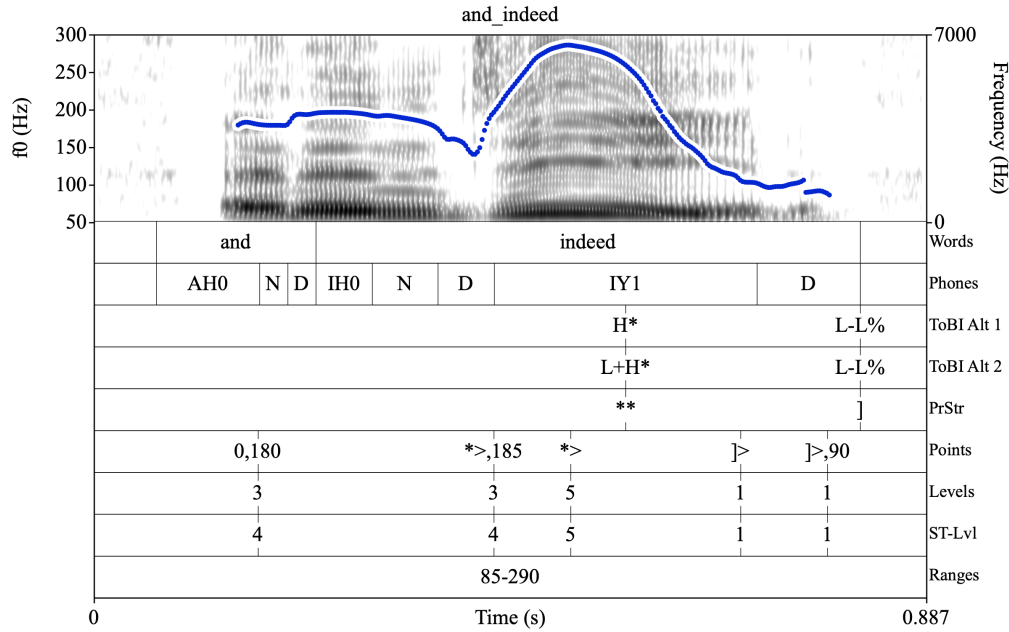


Figure 1.10: MAE_ToBI labellers may disagree or be uncertain about whether a label of L+H* or H* is more appropriate for “indeed”. The different possible labels are indicated on the two “ToBI Alt” tiers.

This disagreement or uncertainty in phonological labels may be because the labellers have different criteria for what constitutes an H* vs. L+H*, or it could be that they differ in their perception of which phonological category is expressed, or both. On the other hand, because Points and PrStr are labelled separately, the PoLaR labels can indicate which aspects of the signal were perceived differently by different labellers. For example, in Figure 1.10, this PoLaR labeller analyzes the lower f0 at the beginning of the vowel as associated with the prominence (using the Advanced *> label on the Points tier), but another labeller might leave this as a 0 label indicating they are not analyzing it as associated with the prominence. This information will be useful in focusing discussions about which phonological label (if any) is most appropriate.

In sum, PoLaR labels can lead to more fruitful discussions about how aspects of the signal map onto categorical labels, making it possible to identify some of the issues with categorical labels related to how different labellers interpret the signal differently (unbeknownst to themselves and/or each other). In addition, if PoLaR labellers want to include MAE_ToBI labelling, the PoLaR labels can be used as a guide. For example, a decision could be made about which labels ought to be H* and which should be L+H*, e.g., by following an algorithm based on the time-alignment of Points with respect to syllable boundaries, and/or pitch-scaling of Levels.

1.2.2 Unlabellable Cases Minimized by PoLaR’s Design

One of the goals of (phonological) grammars is to predict what sorts of expressions are possible and which are impossible. Being able to accurately make such predictions is the marker of a successful phonological analysis, and many prosodists and intonationists have the goal of understanding where the boundaries of well-formedness lie, and thus make proposals to that

end. Some of these proposals constrain systems of prosodic annotation, sometimes to the extent that the annotation system is unable to transcribe a signal that does not conform to the proposed grammar. This issue —the existence of what we call “unlabellable” cases— is one that such systems cannot be used to address, without serious modification.

On the other hand, PoLaR is not tied to any particular phonological grammar, and thus there are no unlabellable cases. Though there are no unlabellable utterances, there is a system for annotation that a researcher can use to describe these otherwise-unlabellable utterances. The benefits of this are rather clear, and below we provide a few examples of cases where PoLaR can be used where other phonologically-tied annotation systems cannot be.

As a first example, in exploratory studies on un(der)documented languages and varieties, most or all utterances are “unlabellable” in the sense that one would need some sort of prosodic theory of the language/variety to start using a phonologically-defined labelling system. In many cases, using an existing annotation system may not be appropriate (because of grammatical dissimilarity among the languages/varieties), and creation of a new phonological annotation system may not be immediately feasible. PoLaR might be particularly useful in such cases, to aid in finding describable and analyzable patterns. As noted above, for many PoLaR labels (especially Basic labels), minimally-trained annotators can systematically label essential aspects of prosody, even if they are not a speaker of the language/variety being studied. For the PoLaR labels that do require speaker intuitions (e.g., PrStr labels), researchers could use experimental methodologies (such as RPT tasks) to crowdsource judgments from speakers of the language.

After PoLaR-annotating speech data from such languages/varieties, researchers in developing phonological analyses can use the patterns in their PoLaR labels to identify aspects of the signal that signal different prosodic categories (e.g., shapes defined by Points/Levels labels that are timed with respect to PrStr labels). For example, Kapia & Brugos (2019) and Brugos & Kapia (2021) have begun to use this methodological division of labor in the exploration of the intonation of Albanian, with PrStr tier annotations informed by both RPT from naïve native speakers and by a trained native speaker of Albanian, and aspects of acoustic phonetic prosody (Points and Ranges) being annotated by a trained labeller who is not a speaker of the language.

In addition to the cases that are “unlabellable” because no phonological labelling system exists (i.e., in early investigations of languages/varieties), phonological labelling systems (by design) also limit what is labellable to utterances whose forms are predicted to be possible by the grammar of the labelling system. In other words, if a phonological labelling system is in any way incomplete or inaccurate in the boundaries it sets for what is prosodically (im)possible, there will necessarily be utterances that are “unlabellable”. (In addition, because some cases are indeed unlabellable, a researcher may find themselves wondering about particular (challenging) examples “*Does this conform to the grammar in my theory, and if I’m not sure, which labels should I use?*”. This can lead to issues of uncertainty or disagreements about labels, of the sort described in the previous section.)

As a concrete example of utterances that are “unlabellable” because their prosodic properties are not predicted possible by the grammar, consider the mainstream U.S. English yes/no questions below:

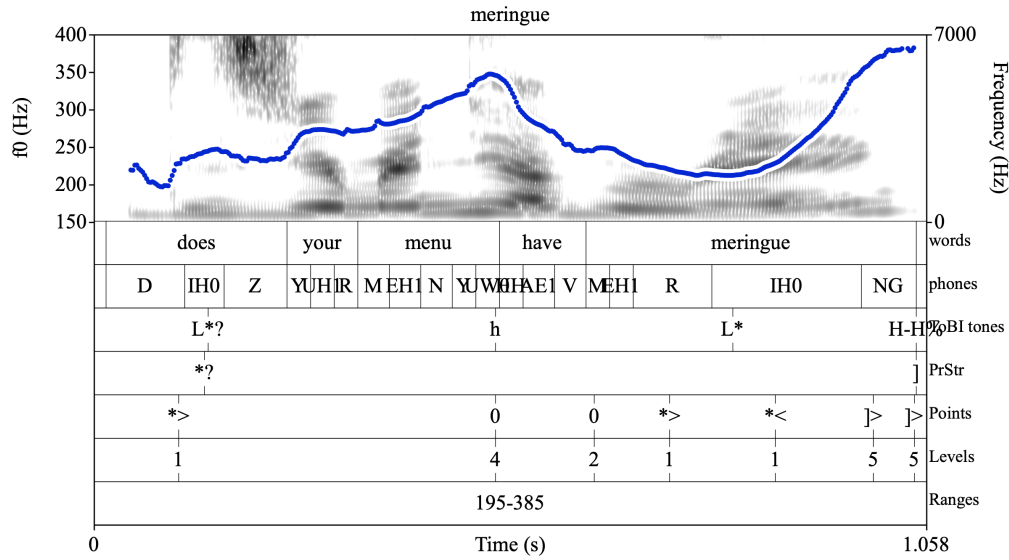


Figure 1.11: A steady rise to a peak in the final (unstressed) vowel of “menu”, which is not perceived as prominent or at a phrase boundary.

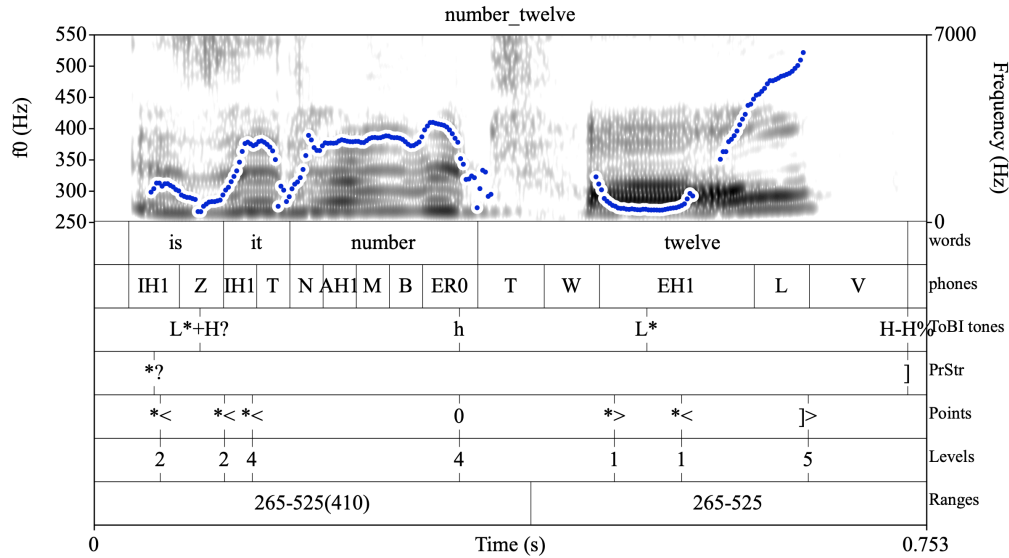


Figure 1.12: A steady high through the final vowel of “number”, which is not perceived as prominent or at a phrase boundary.

As described in Ahn & Zhou 2019, these examples cannot be labelled using MAE_ToBI: the rises are not tied to stressed syllables or prosodic phrase boundaries, but the MAE_ToBI phonology requires that all pitch movements labelled on the Tones tier be associated with one of those two types of objects. As such, they are unlabellable in MAE_ToBI, unless the labelling conventions are broken (e.g., using H* in cases where there is no prominence) or set of labels is augmented (e.g., adding an h label, as used in the figures above). Thus, strictly adhering to MAE_ToBI would require not labelling these pitch movements, making it difficult to keep track of (and subse-

quently analyze) where these unpredicted highs occur. PoLaR labels, on the other hand, could be used for this utterance just as easily as it is used for any other utterance: all pitch movements are labelled with the Points and Levels tiers, and can be used for description and analysis.

In sum, using PoLaR to label otherwise unlabellable examples will encourage researchers to analyze patterns they might not otherwise (be able to) track. When there is no established prosodic grammar for a language/variety, PoLaR can be used to help build one. Additionally, using PoLaR alongside a phonological labelling system (when available) may help refine our understanding of the prosodic phonology (and may provide useful evidence for discussions about updating phonological annotation systems).

Bibliography

- Ahn, Byron & Zhou, Z.L. 2019. Is this in the phonology? examining the intonational phonetics-phonology interface with American English polar questions. In Calhoun, Sasha & Escudero, Paola & Tabain, Marija & Warren, Paul (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*, 2450–2454. Canberra, Australia: Australasian Speech Science and Technology Association Inc.
- Albert, Aviad & Cangemi, Francesco & Grice, Martine. 2018. Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. In *Speech Prosody 2018*. 804–808. doi:10.21437/SpeechProsody.2018-162
- Barnes, Jonathan & Brugos, Alejna & Veilleux, Nanette & Hufnagel, Stefanie Shattuck. 2014. Segmental influences on the perception of pitch accent scaling in English. In Campbell, Nick & Gibbon, Dafydd & Hirst, Daniel (eds.), *Proceedings of speech prosody 7*, 1125–1129. Dublin: International Speech Communication Association.
- Barnes, Jonathan & Veilleux, Nanette & Brugos, Alejna & Shattuck-Hufnagel, Stefanie. 2010. Turning points, tonal targets, and the English L- phrase accent. *Language and Cognitive Processes* 25. 982–1023.
- Barnes, Jonathan & Veilleux, Nanette & Brugos, Alejna & Shattuck-Hufnagel, Stefanie. 2012. Tonal center of gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology* 3(2). 337–383. doi:10.1515/lp-2012-0017
- Beckman, Mary E. 1993. Modeling the production of prosody. In House, David & Touati, Paul (eds.), *Proceedings of ESCA workshop on prosody*, 258–263. Lund, Sweden: Department of Linguistics and Phonetics, University of Lund. http://www.isca-speech.org/archive_open/prosody_93.
- 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.
- Bruce, Gösta & Granström, Björn & Gustafson, Kjell & Horne, Merle & House, David & Touati, Paul. 1997. On the analysis of prosody in interaction. In Sagisaka, Yoshinori & Campbell, Nick & Higuchi, Norio (eds.), *Computing prosody: Computational models for processing spontaneous speech*, 43–59. New York, NY: Springer. doi:10.1007/978-1-4612-2258-3_4
- 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 & Kapia, Enkeleida. 2021. Prosodic and non-prosodic cues to prominence and boundaries: Evidence from an RPT study in Albanian. Poster presentation at the 1st International Conference on Tone and Intonation (TAI).

- Burdin, Rachel Steindel & Holliday, Nicole & Reed, Paul E. 2018. Rising above the standard: Variation in L+H* contour use across 5 varieties of American English. In *Proc. 9th international conference on speech prosody 2018*. 354–358. doi:10.21437/SpeechProsody.2018-72
- Carlson, Rolf & Hirschberg, Julia & Swerts, Marc. 2005. Cues to upcoming Swedish prosodic boundaries: Subjective judgment studies and acoustic correlates. *Speech Communication* 46(3-4). 326–333. doi:10.1016/j.specom.2005.02.013
- d'Alessandro, Christophe & Mertens, Piet. 1995. Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language* 9(3). 257–288. doi:10.1006/csla.1995.0013
- Dehé, Nicole. 2009. Clausal parentheticals, intonational phrasing, and prosodic theory. *Journal of Linguistics* 45(3).
- 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.
- d'Imperio, Mariapaola. 2000. *The role of perception in defining tonal targets and their alignment*: The Ohio State University dissertation.
- Fagyal, Zsuzsanna. 2002. Tonal template for background information: the scaling of pitch in utterance-medial parentheticals in French. In *Speech Prosody 2002*, ISCA Archive.
- Féry, Caroline & Truckenbrodt, Hubert. 2005. Sisterhood and tonal scaling. *Studia Linguistica* 59(2-3). 223–243.
- Gasser, Emily & Ahn, Byron & Napoli, Donna Jo & Zhou, Z.L. 2019. Production, perception, and communicative goals of American newscaster speech. *Language in Society* 48(2). 233–259. doi:10.1017/S0047404518001392
- Hansson, Petra. 2003. *Prosodic phrasing in spontaneous Swedish*: Lund University dissertation.
- Hirschberg, Julia. 2004. Pragmatics and intonation. In Horn, Laurence R. & Ward, Gregory (eds.), *The handbook of pragmatics*, 515–537. Malden, MA: Blackwell.
- Hirschberg, Julia & Pierrehumbert, Janet. 1986. The intonational structuring of discourse. In *24th annual meeting of the association for computational linguistics*. 136–144.
- Hirschberg, Julia & Ward, Gregory. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics* 20(2). 241–251.
- Hirst, Daniel. 1993. Peak, boundary and cohesion characteristics of prosodic grouping. In House, David & Touati, Paul (eds.), *Proceedings of ESCA workshop on prosody*, 32–37. Lund, Sweden: Department of Linguistics and Phonetics, University of Lund. https://www.isca-speech.org/archive_open/prosody_93/pro3_032.html.
- Holliday, Nicole. 2021. “the switch up is real”: Examining prosodic correlates of style shifting on Twitter. Paper presented at NWAV 49.
- Kapia, Enkeleida & Brugos, Alejna. 2019. Perception of prosodic correlates of topic and focus in albanian. Poster at Intonational Phonology of Typologically Rare or Understudied Languages, (ICPhS 2019 Satellite Workshop).
- Kentner, Gerrit & Féry, Caroline. 2013. A new approach to prosodic grouping. *The Linguistic Review* 2013 30. 277–311.

- 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. 1992. Compound prosodic domains. *Occasional Papers of the Department of Linguistics, University of Edinburgh*.
- Lin, Hsin-Yi & Fon, Janice. 2011. The role of pitch reset in perception at discourse boundaries. In Lee, Wai Sum & Zee, Eric (eds.), *Proceedings of the XVIIth International Congress of Phonetic Sciences*, 1242–1245. Hong Kong: City University of Hong Kong.
- Niebuhr, Oliver & D’Imperio, Mariapaola & Fivela, Barbara Gili & Cangemi, Francesco. 2011. Are there “shapers” and “aligners”? individual differences in signalling pitch accent category. In Lee, Wai Sum & Zee, Eric (eds.), *Proceedings of the XVIIth International Congress of Phonetic Sciences*, Hong Kong: City University of Hong Kong.
- Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*: MIT dissertation.
- Price, Patti & Ostendorf, Mari & Shattuck-Hufnagel, Stefanie & Fong, Cynthia. 1991. The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America* 90(6). 2956–2970. doi:10.1121/1.401770
- Rett, Jessica & Sturman, Beth. To appear. Prosodically marked mirativity. *WCCFL 38: Proceedings of the 38th West Coast Conference on Formal Linguistics*.
- Sicoli, Mark & Stivers, Tanya & Enfield, N.J. & Levinson, Stephen C. 2015. Marked initial pitch in questions signals marked communicative function. *Language and Speech* 58(2). 204–223. doi:10.1177/0023830914529247
- Sturman, Bethany. 2019. The emphatic juncture: A novel use of the IP boundary in English. 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.
- Thomas, Erik. 2007. Phonological and phonetic characteristics of African American Vernacular English. *Language and Linguistics Compass* 1(5). 450–475. doi:10.1111/j.1749-818x.2007.00029.x
- van den Berg, Rob & Gussenhoven, Carlos & Rietveld, Toni. 1992. Downstep in Dutch: implications for a model. In Docherty, Gerard J. & Ladd, D. Robert (eds.), *Papers in laboratory phonology II: Gesture, segment, prosody*, Cambridge: Cambridge University Press.
- Wichmann, Anne. 2000. *Intonation in text and discourse: Beginnings, middles and ends*. Pearson Education Limited.
- Wolfram, Walt & Thomas, Erik. 2002. *The development of african american english*. Blackwell Publishers.
- Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics* 27. 55–105.
- Xu, Yi & Xu, Ching X. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33(2). 159–197.