ANNOTATING PROSODY WITH POLAR: CONVENTIONS FOR A DECOMPOSITIONAL ANNOTATION SYSTEM

Byron Ahn[1], Nanette Veilleux[2], and Stefanie Shattuck-Hufnagel[3]

[1]Princeton University, [2]Simmons College, [3]MIT
bta@princeton.edu, veilleux@simmons.edu, sshuf@mit.edu

ABSTRACT

An utterance's intonational characteristics vary according to linguistic meaning; even controlling for meaning, they may vary across speakers, and even across utterances for a given speaker. Phonological annotation systems typically bundle together disparate characteristics, such as f0 scaling, alignment, and prominence (e.g., labels like H*). This can make it difficult to document variation, and also to determine which aspects of the signal (i) result from the phonology-phonetics interface, (ii) vary according to abstract linguistic features, (iii) depend on dialect, social context, or emotional state, or (iv) may be simply noise. This has motivated a new annotation system: Points, Levels, and Ranges (PoLaR). PoLaR takes inspiration from (and can be used alongside) existing theoretical (AM theory) and transcriptional (IPO, IViE, RPT) systems, but is neither a phonological/cognitive model, nor an acoustic/physical model. It isolates individual prosodic characteristics, using labels that transparently correspond to aspects of the acoustics and/or native speaker perception.

Keywords: Annotation, Intonation, Prosody, Acoustic cues, Meaning

1. INTRODUCTION

This paper introduces PoLaR, an annotation framework designed to annotate salient aspects of the intonational acoustics, with an eye towards exploring the relationship between variation and intonational phonology. The motivating goals for this new annotation system are to (i) capture more of the systematic variation that is often observed and (ii) minimize difficulty in labeling, by decomposing a phonological category into its component parts.

Existing frameworks for annotating intonation in English, while useful for many purposes, are on their own either ill-suited for relating individual acoustic cues to phonological categories or even ill-suited for simply annotating acoustics at all. In particular, some are too phonetically-driven to relate surface forms to phonological models (e.g., American structuralism [12], IPO [9]). Others aim to annotate intonational variation, but the labelling system is more like broad phonetic transcription, necessitating commitments to particular phonological models (e.g., IViE [7], IPrA [10]), which then requires the labellers to be familiar with such phonological models and take on their assumptions about the phonological categories. For yet others, the same sorts of issues are magnified, because the labeling systems are defined in abstract/categorical terms that do not allow labeling of fine-grained intonational variation (e.g., ToBI [3], RPT [4]).

On the other hand, PoLaR isolates particular intonational characteristics (e.g., pitch range) on individual tiers, whose labels are time-aligned to recordings. In other transcription systems based on the AM tradition [11], labels are intended to stand in for a constellation of characteristics: e.g., H* may represent some if not all of the following characteristics: an f0 value that is high in the local pitch range, a turning point in the f0 contour, and perceived abstract post-lexical (i.e., phrase-level) prominence. Labels that bundle an extensive inventory of properties in this way are absent from PoLaR. Instead, its phonological labels simply capture the presence of a prominence or a boundary (similar to RPT), on a phonological tier.

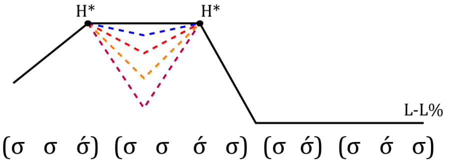
PoLaR's decompositional annotations also specify the acoustic characteristics of pitch range, scaled pitch levels within that range, and f0 turning points, on different tiers. By design, PoLaR also serves as an exploratory tool to investigate the acoustic cues to meaningful categories, by allowing for deeper analysis of the systematicities in phonetic/phonological relationships between these characteristics. While some such relationships have been discussed in decades of work in the AM tradition, many past labelling systems (due to the intentional design features of these systems) have expressly not captured phonetic variation associated with phonological categories. PoLaR does not replace phonological and broad phonetic transcriptions of other annotation systems, but can supplement them. For example, it enables comparison of items labeled L+H* versus H* in a ToBI transcription, to see what characterizes these different labels. Lastly, PoLaR allows researchers to investigate the relationship between particular acoustic characteristics of intonation and other (semantic, pragmatic, sociolinguistic, and paralinguistic) aspects of the speech context.

2. MOTIVATION

As noted above, the development of PoLaR was motivated by two main goals: (i) to isolate phonetic cues to supplement more phonological labels that encompass multiple intonational dimensions, and (ii) to allow annotators to use transparent and easy-to-use labels for these cues. The former will be useful for considering what role such patterns might play in the grammar and/or in the communication process.

As an example, consider the different idealized pitch contours in Figure 1 that represents a familiar variation across a range of realizations of 'the same' contour, with some degree of f0 sagging. At least some subset of these (if not all) may be assigned the same phonological representation (e.g., in MAE_ToBI: H* H* L-L%). What (if anything) conditions this sagging, and its depth? Perhaps the triggers are phonological, and this is a case of phonologically conditioned allophony. Or, perhaps it is tied to linguistic meaning (e.g., stance [6]). Perhaps the variation is meaningful throughout a linguistic community, tied to individual speakers or part of non-communicative variation. Until such variation is systematically annotated (which PoLaR allows for), exploring the conditioning factor(s) is a serious challenge.

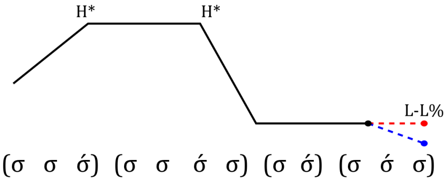
Figure 1: F0 variation between pitch peaks



Alternatively, perhaps the f0 tracks in Figure 1 realize more than one phonological representation: some might be a phonological H* H* with a phonetic sag, while others are a phonological H* L+H* sequence. This would raise questions concerning whether there is a crossover point corresponding to categorical boundaries, or whether the categories' realizations are more overlapping in their distributions (as a case of (incomplete) neutralization). Without a system like PoLaR to capture the acoustic variation, investigating these issues is more challenging.

As another example, do the two f0 contours at the end of Figure 2 share the same meaningful distribution? If not, systems which label them the same (e.g., L-L%) might require adjustment to the set of phonological labels. It would be difficult to make this determination, without annotating the two as phonetically different (which PoLaR allows).

Figure 2: F0 variation in phrase-final falls



Because of the lack of clarity in these issues, both novice and experienced labellers regularly encounter f0 contours which may raise questions about which categorical label (if any) is appropriate to use in intonational annotation. Similarly, various researchers have also noticed that some of these difficult-to-label contours are actually potential candidates for signaling meaningful differences, and have thus innovated ad hoc annotation solutions (e.g., [2], [5]); these ad hoc solutions unfortunately prevent easy comparison across groups and projects. The PoLaR system is designed to fill this need, supplementing a purely phonological approach (e.g., that of ToBI) with a set of phonetic labels that track the acoustic signal more closely, in order to explore which aspects of variation in that signal might be systematically related to meaning differences.

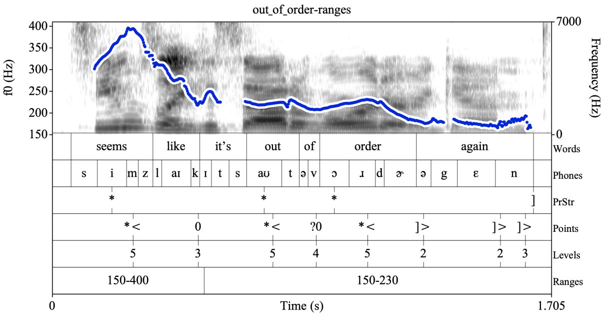
3. SKETCHING OUT THE POLAR ANNOTATION SYSTEM

To meet these needs, PoLaR was developed with more transparent labels for individual intonational characteristics. A greater number of tiers are used to disentangle intonational features and characteristics, allowing labellers with a variety of phonological assumptions to agree on how to label each tier. An example annotated both with PoLaR labels is given in Figure 3.

PoLaR builds on the AM tradition, adopting general views on prominence, phrasing, and a relationship between these and certain tonal values in the f0 contour. However, unlike systems that provide phonological or broad phonetic transcriptions, PoLaR labels do not depend on an understanding of what is prosodically (im)possible in the language's grammar.

Instead, the PoLaR system consists of (at least) four tiers: a Prosodic Structure tier that captures some broad phonological categories, and three

Figure 3: Example with PoLaR annotation



acoustic-based tiers: the Points, Ranges, and Levels tiers. The first three of these can be labelled completely independently of one another. (At the same time, optional labels allow labellers to indicate relationships between the tiers, as well.) The tiers and their labels are very briefly described below; further details are provided in the PoLaR labelling guidelines ([1]).

The Prosodic Structure tier annotates the abstract prosodic events of prominence and phrasing, marking only their presence with the basic labels `*` and `]`. Additional (optional) labels can capture distinctions in types of prominence/phrasing, if the labeller is comfortable making such distinctions. No details about their acoustic realization are labelled, similar to RPT [4]. See Table 1 for a brief summary of this tier’s labels.

The (Pitch) Points tier captures turning points in the f0 contour, without any information about the (relative) pitch value/height. The default label is “`0`” which identifies the f0 turning point as an acoustic event, without any commitment to its (potential) phonological status. Optional labels can encode the labeller’s analysis of any relationship to the phonological events on the Prosodic Structure tier labels (e.g., `*>` means turning point is associated with labeled prominence to the right). See Table 2.

The Range (Domains) tier indicates the local f0 speaking range, defining approximately where “high” and “low” are, for the annotated portion of the utterance. (This is distinct from the speaker’s comfortable pitch range, or the range of pitch they use in the entire conversation.) Human annotation is necessary to determine whether a speaker has changed their range as they move through an utterance. This tier has the least corresponding literature or research history to guide annotators, but it permeates decisions in many other annotation systems (e.g., deciding to label a pitch accent as `H*`, even if its f0 is similar to other `L*` accents). The maximum and minimum f0 are rounded up/down from the software-measured f0 max/min to create a wider range. Intervals of time where f0 is not reliable enough for inferring a range boundary can be labeled `NA`. (See Table 3.)

The (Scaled) Levels tier annotates the relative and local pitch scaling of a Points label, based on the local pitch range. As the value is simply applied by dividing the range into 5 intervals and comparing the f0 value at the time to which the Point is aligned, these labels can be automatically populated from the Ranges and Points tiers. Each label on this tier is given a value, 1–5, for how high/low the event is in the local pitch range (similar to [12] and [8]). A level of 1 represents a Point whose f0 value in the speaker’s lowest quintile in the local pitch range, while a level of 5 represents one in the speaker’s highest quintile.

4. CONCLUSIONS

PoLaR is not conceived of as a grammar: what is labelled may or may not be reflective of the abstract systems that deal in intonation contrasts. Instead of providing such a theory, PoLaR is a methodological tool intended for use by many different types of researchers, including all types of descriptivists and theorists. As such, not all of its labels correspond directly to some aspect of the signal, nor do they necessarily correspond directly to some abstract categories in the phonology: PoLaR representations are intermediary in that they are not exclusively reflective of either the physical or the phonological.

Instead, PoLaR was designed to make minimal phonological assumptions while capturing a maximal amount of the relevant acoustics-phonetics. The phonological aspects of PoLaR only assume that there are phonological events that native speakers have intuitions about, supported by the intonational signal (a core feature of all AM-based labelling systems). In addition, acoustic annotations refer only to aspects of the signal presumed relevant for intonational contrasts. As such, PoLaR provides a means of transparently annotating individual characteristics of the intonational signal, allowing phonological/phonetic labelling of more variation as well as of tokens that are challenging for existing phonological models to capture. This will facilitate the exploration of new areas in the domain of intonation and prosody, including: the intonational phonetics-phonology interface, the relationship between prosody and meaning, intra-linguistic prosodic variation, and dialects (or even languages) with little to no existing understanding of the prosodic system.

PoLaR, as it currently stands, focuses on prosody

Table 1: Some Prosodic Structure ('PrStr') tier labels for English. The shaded rows show the core labels for this tier; other labels are considered optional, for use by more practiced labellers who feel able to include finer details about prominence and phrasing.

Label	Phonological object	Time-aligned with:
*	Prominence	The temporal mid-point of a vowel that has post-lexical prominence
*?	Possible Prominence	The temporal mid-point of a vowel when the labeller is uncertain of whether there is post-lexical prominence
**	Highest Prominence	The temporal mid-point of a vowel that has uniquely strong post-lexical prominence
]	Phrase Right Edge	The right edge of the final word of a phrase
]?	Phrase Right Edge	The right edge of the final word when the labeller is uncertain of whether there is a phrase boundary
]]	Large Phrase Right Edge	The right edge of the final word of a relatively larger phrase

Table 2: Some Points tier labels for English, some of which relate the f0 contour and the Prosodic Structure tier. The 0 label (shaded) is the only necessary label. Other labels are considered optional, for use by labellers who have intuitions about the relationship between f0 and prosodic structure.

Label	Relevant Prosodic Structure object	This f0 turning point is time-aligned ______
0	None	N/A
*>	Prominence *, *?, or **	before the relevant * on the PrStr tier
*<		after the relevant * on the PrStr tier
*@		with the relevant * on the PrStr tier
]>	Phrase boundary],]?, or]]	before the relevant] on the PrStr tier
]<		after the relevant] on the PrStr tier
]@		with the relevant] on the PrStr tier

Table 3: Labels for intervals on the Ranges tier. [*min*] and [*max*] are to be replaced by appropriate numerical values, without surrounding brackets. Most intervals on this tier will be given a label from the shaded row.

Label	Meaning:
[*min*]-[*max*]	[*min*] or [*max*] is the local pitch minimum or maximum, respectively, in Hertz, rounded down/up to nearest multiple of 5
X-[*max*]	"X" represents a pitch minimum that the labeller cannot determine
[*min*]-X	"X" represents a pitch maximum that the labeller cannot determine
NA	"NA" indicates a stretch of unreliable pitch tracking, where the pitch range minimum and maximum cannot be inferred

as it relates to English intonation. Beyond this, PoLaR is intended as an extendable framework. Additional tiers or labels can be added, with no impact on the existing ones. It may be particularly useful to add a Voice Quality tier, or labels in the PrStr tier for additional phrase boundary types. In this way, our hope is that PoLaR's flexibility and utility will contribute to deepening the field's understanding of prosody, at both empirical and theoretical levels.

5. REFERENCES

[1] Ahn, B., Veilleux, N., Shattuck-Hufnagel, S., Brugos, A. 2019. PoLaR annotation guidelines. Available at http://doi.org/10.17605/OSF.IO/USBX5.

[2] Ahn, B., Zhou, Z. 2018. Spurious pitch movements in American English polar questions. Presented at Experimental and Theoretical Advances in Prosody 4. Available at http://doi.org/10.17605/OSF.IO/4R79Q.

[3] Beckman, M. E., Hirschberg, J., Shattuck-Hufnagel, S. 2005. The original ToBI system and the evolution of the ToBI framework. In: *Prosodic Typology: The Phonology of Intonation and Phrasing*.

[4] Cole, J., Mahrt, T., Roy, J. 2017. Crowd-sourcing

prosodic annotation. *Computer Speech and Language* 45, 300–325.

[5] Dehé, N. 2018. The prosody of rhetorical questions. *NELS 48: The Proceedings of the Forty-Eighth Annual Meeting of the North East Linguistic Society* 1, 173–192.

[6] Freeman, V. 2015. Prosodic features of stance acts. *The Journal of the Acoustical Society of America* 138(3), 1838.

[7] Grabe, E., Post, B., Nolan, F. 2001. Modelling intonational variation in English: the IViE system. In: *Proceedings of Prosody 2000*. Poznan, Poland: Adam Mickiewicz University 51–58.

[8] Grabowski, E., McPherson, L. 2018. ATLAS (Automated Tone Level Annotation System): A tonologist's and documentarian's toolkit. *Tonal Aspects of Language 2018*.

[9] 't Hart, J., Collier, R., Cohen, A. 1990. *A perceptual study of intonation: an experimental-phonetic approach to speech melody*. New York: Cambridge University Press.

[10] Hualde, J. I., Prieto, P. 2016. Towards an international prosodic alphabet (IPrA). *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1), 5.

[11] Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation*. PhD thesis MIT.

[12] Pike, K. L. 1945. *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.