An overview of prosodic transcription systems: A comparison of the "Tones and Break Indices: ToBI" and " Rhythm and Pitch: RaP"

Nayereh Joodi[1], Golnaz Modarresi Ghavami[2]
Laura Dilley[3]

1. Introduction

The selection of the appropriate labeling system in any prosodic study depends on the research purpose. In the current research, we have reviewed the labeling system known as Tones and Break Indices (ToBI) (Pierrehumbert & Hirschberg, 1990) and its alternative labeling systems including Rhythm and Pitch (RaP) (Dilley, 2005; Dilley & Brown, 2005; Dilley et al., 2006). The problems of the ToBI system were summarized and presented. Furthermore, a review of the studies conducted on intonation in Persian using the ToBI system within the framework of Auto-segmental Metrical theory (AM) showed that the global problems of this system is also observable in its application for the analysis of Persian intonation patterns (e.g. Eslami, 2005; Sadeghi, 2018).

Originally, the main goal of ToBI was to provide a standard transcription tool for labeling intonational features, including prominence patterns and prosodic structure of an utterance so that different users with different working fields could use and interpret each other's linguistic data. In the ToBI transcription system, L and H represent low and high tones, respectively. The diacritic * represents pitch accent, and % represents boundary tones (Beckman & Elam, 1997). This system was initially designed for transcribing the intonation and prosodic structure of English utterances (Silverman et al., 1992; Beckman & Hirschberg, 1994; Beckman & Ayers Elam, 1997; Beckman, Hirschberg, and Shattuck- Hufnagel 2005), as well as a few typologically different languages— for example, GToBI for German (Grice & Benzmüller, 1995), K- ToBI for Korean (Beckman & Jun, 1996; Jun 2000), and J_ToBI for Japanese (Venditti, 1997), and Persian (Eslami,2005). Jun (2022), Ladd (2022), and Dilley and Breen (2022) have identified the shortcomings and problems of the ToBI phonetic labeling system to create an International Prosodic Alphabet (IPrA) (Hualde & Prieto, 2016).

---

[1] Ph.D. Candidate of Linguistics, Allameh Tabataba'i University; nirvanajoodi1400@gmail.com; joodi_n@atu.ac.ir
[2] Associate professor of Linguistics, Allameh Tabataba'i University (corresponding author); modarresighavami@atu.ac.ir
[3] Associate professor, Dept. of Communicative Sciences and Disorders,College of Communication Arts & Sciences, Michigan State University; ldilley@msu.edu

The Rhythm and Pitch (RaP) system based on enhanced Auto-segmental Metrical theory (AM$^+$) was proposed by Dilley and her colleagues (Dilley, 2005; Dilley & Brown, 2005; Dilley et al., 2006; Dilley & Breen, 2012) to overcome the difficulties of ToBI in showing variations and gradation of the categories and to emphasize the importance of distinguishing rhythmic or metrical prominence from pitch prominence. In this system, pitch information is labeled as three tonal targets (H, L, E) and compared to the previous pitch pattern (higher, lower, or equal to it) in the speech signal. Therefore, labels in RaP have a phonetic representation. Metrical prominence (at three levels of strong, weak, and none) and prosodic structure (at two levels, intonational phrase (IP) and intermediate phrase (ip)) are labeled in the rhythm layer. Although RaP was presented as a "method of transcribing rhythm and related pitch in English" (Dilley and Brown, 2005 p, 2), the concepts and principles of this system can be applied to other languages.

2. Materials and methods

In the current research, we investigated the efficiency of the RaP system in labeling Persian samples and compared this labelling with ToBI labels. A female speaker of Persian reproduced one sample sentence of Eslami (2005) (/man cetab xaridam/ 'I bought a book') with an unmarked prosody and with a focus on each component of the sentence. This sample was recorded and subsequently labeled in PRAAT (version 6.1.48) using Persian-ToBI and RaP so that the labels of these two systems could be compared with each other (Figure 1).

Figure   **1**

Unmarked reading: /man cetab xaridam/ 'I bought a book'. To compare the labels of both systems, this sentence was recorded and labeled in PRAAT (version 6.1.48) using Persian-ToBI and RaP.
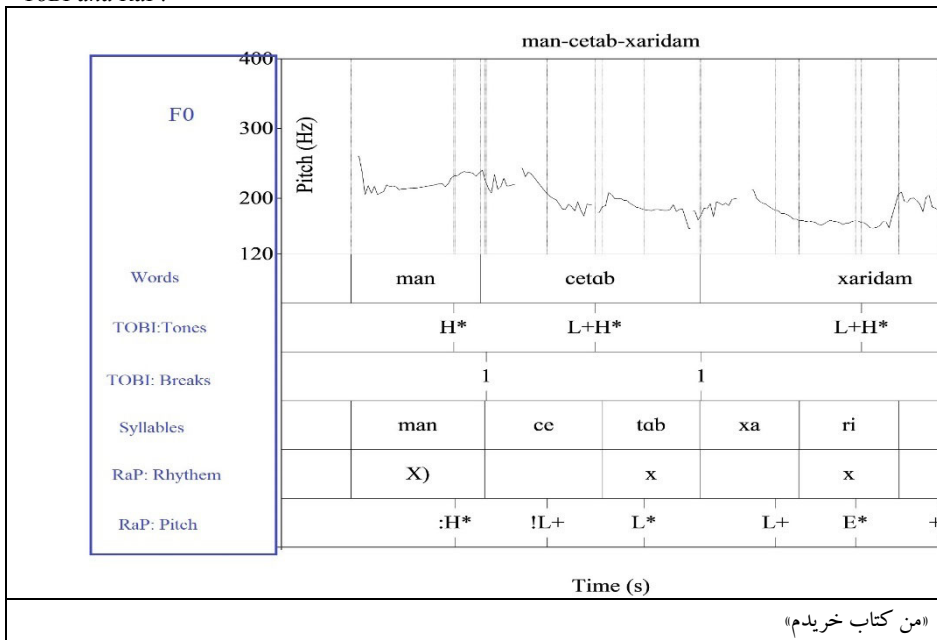
3. Results and discussion

In PersianToBI, it has been shown that the default pitch pattern of words is L+H* (Eslami, 2005), which can have different allophones (H*) in different contexts. Due to its phonological nature, this system does not reflect phonetic differences between sentences. Therefore, in unmarked reading, all constituents of the sentence in ToBI are labeled as L+H*. In focal reading of the intended sentence, ToBI assigns the same labels to the focused words. In general, pitch is reduced after focused words in the focal sentence (Taheri-Ardali,et al., 2014). This intonational pattern is labelled as L-L% in the ToBI System. In ToBI, the rhythm is not shown, while many studies have shown that rhythm plays an important role in prosodic typology, sentence comprehension, and disambiguation of garden paths (Breen et al. 2014; Dilley et al. 2010; Dilley & McAuley 2008; Morrill et al. 2014; Dilley & Breen, 2022; Jun, 2014; Tilsen & Arvaniti, 2013).

In RaP, differences in the tonal pattern between the unmarked reading and readings with focus on different words in the sentence are labelled differently according to the observed pitch contour and the perception of the researcher. RaP shows rhythm in the rhythm tier and uses X, x on accented and unaccented syllables. In RaP, the most prominent stressed syllable and the highest level of the speaker's pitch range are marked with >H* (the focal word in each sentence) and is labeled X in the rhythm tier.

4. Conclusion

A Comparison of the two labelling systems demonstrates the phonetic clarity of RaP. The following are some of RaP's strengths: a) RaP is easier to learn and use compared to ToBI, b) RaP labels correlate uniformly with phonetic and perceptual features based on listeners' perception and phonetic information, c) each syllable receives a label related to starred or unstarred tones in RaP (Arvaniti et al., 1998; Ladd et al., 2000; Dilley et al., 2005), d) using this system, rhythmic information is separated from pitch information, allowing researchers to identify and label different levels of salience.