A proposal for labelling prosodic disfluencies in ToBI

Alejna Brugos, Boston University, and Stefanie Shattuck-Hufnagel, MIT

The ToBI transcription system (Pitrelli et al. 1994, Silverman et al. 1992) was developed during a series of 1990s workshops (Beckman et al. 2005) as an initial draft of a system for labelling intonational constituents and prominences, reflecting the phonological theory embodied in Pierrehumbert (1984), Beckman and Pierrehumbert (1986) and related works. It was envisioned that a common transcription system would enable a) sharing of prosodically labelled databases (Ostendorf et al. 1985), b) cross-language comparison of intonational prosody (Jun 2005) and c) identification of labelling issues not yet well-handled and requiring further development.

Pierrehumbert's and Beckman's theory was developed on the basis of utterances with contrastive intonational elements, often produced in citation-form read laboratory speech. In the intervening years there has been a shift of interest toward both the prosody and the segmental phonetics and phonology of spontaneously produced speech (in corpora such as Switchboard, Callhome, Buckeye). Such speech is characterized by a large number of disfluencies, involving less-than-fluent prosody on the one hand, and mistakes in words or speech sounds in the other. The original ToBI system included labels for such disfluencies (known as 'p-breaks', i.e. 1p, 2p, 3p, related to Break Indices 1, 2 and 3). These labels were based on disfluency types identified by Nakatani & Shriberg (1993), but they were designed to flag disfluencies only roughly, leaving aside much of the information that might shed light on how these events interact with prosodic structure. In addition, the p-break categories are not always intuitive, so that the high rate of disfluencies in spontaneous speech often absorbs a disproportional amount of a labeller's time.

The proposal described here builds on existing ToBI conventions for the labelling of disfluencies, using a simplified version of the approach to disfluency labelling developed by Arbisi-Kelm (2006, 2010) for stuttered speech, and incorporates aspects of disfluency annotation systems developed for other languages, such as the disfluency-related extensions of X-JToBI for Japanese (Maekawa et al. 2002), and those of Rodriguez et al (2001) for Spanish. It introduces explicit disfluency-related labels into the Break Index tier to cover the majority of types of disfluent events described by Nakatani and Shriberg, while unpacking the portmanteau of disfluency types covered by the original p-break labels into separate recombinable labels for i) filled pauses, ii) disfluent (segmental) prolongation, iii) silent pauses and iv) cut-off words (see Table 1). Also under development are explicit labels for word or segment errors, and restarts. This system for labelling prosodic disfluencies separately from 'speech errors' (such as word and sound sequencing errors), with a more fine-grained categorization of the prosodic disfluency type, will enable the separate study of these two phenomena, as well as of their interaction. Thus, it has the potential (along with the ALT tier proposed in Brugos et al. 2008, Brugos et al. 2006) to materially increase the speed, reliability and convenience of the ToBI labelling process. Continuing development of this proposal will welcome input from researchers who have also been grappling with these issues.

Table 1: A partial table of disfluency phenomena and proposed labels, to be used with the ToBI Breaks tier.

Phenomenon	Symbol	Description
prolongation	pr	abnormal and/or incongruous prolongation of a segment within a word
cut	c	a partially completed word
disfluent pause	ps, psw	abnormal and/or incongruous pause between (ps) or within (psw) words
silence	s	end of a silence (whether disfluent-sounding or not)
filler	f	filled pause, filler words or segments (e.g. *um, uh, mm*)
error	e	mispronunciation or wrong word
restart word	rs	restarting of a segment, syllable, word, after a word has been cut off
restart phrase	%r	start of a new phrase after a previous phrase was not finished

References:
Arbisi-Kelm, T. (2006), *An Intonational Analysis of Disfluency Patterns in Stuttering*. Ph.D. thesis, UCLA.
Arbisi-Kelm, T. (2010), Intonation structure and disfluency detection in stuttering, In C. Fougeron, B. Kühnert, M. D'Imperio, N. Vallé (eds.) *Laboratory Phonology 10*. Berlin: De Gruyter Mouton, 405-432
Beckman, M. and J. Pierrehumbert (1986), Intonational Structure in Japanese and English, *Phonology Yearbook III*, 15-70.
Beckman, M., Hirschberg, J. and Shattuck-Hufnagel, S. (2005), The Original ToBI System and the Evolution of the ToBI Framework. In Jun, S.A. (Ed), *Prosodic Typology*. Oxford: Oxford University Press, 9-54.
Brugos, A., Shattuck-Hufnagel, S. & Veilleux, N. (2006) MIT OCW 6.911: Transcribing Prosodic Structure of Spoken Utterances with ToBI. Online course: http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/
Brugos, A., Veilleux, N., Breen, M., Shattuck-Hufnagel, S. (2008). The Alternatives (Alt) Tier for ToBI: Advantages of Capturing Prosodic Ambiguity, in: Proceedings of Speech Prosody.
Jun, S.-A. (2005), (Ed), *Prosodic Typology*. Oxford: Oxford University Press.
Ostendorf, M., P. J. Price and S. Shattuck-Hufnagel (1985), The Boston University Radio News Corpus, *Boston University Technical Report No. ECS-95-001*, March 1995
Maekawa, K., Kikuchi, H., Igarashi, Y. & Venditti, J. (2002) X-JToBI: An extended J-ToBI for spontaneous speech. In *Proceedings of the 2002 International Conference on Spoken Language Processing (ICSLP)*. Denver, Colorado.
Nakatani, C.H. & Shriberg, E.E. (1993). Proposal for labeling disfluencies in ToBI. Paper presented at the Third ToBI Labeling Workshop, Ohio State University.
Pierrehumbert, J. (1980), *The Phonology and Phonetics of English Intonation*. Ph.D. thesis, MIT.
Pitrelli, J., Beckman, M.., Hirschberg, J. (1994): Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *ICSLP-1994*, 123-126.
Rodríguez, L.J., Torres, I. , Varona, A. (2001) Annotation and analysis of disfluencies in a spontaneous speech corpus in Spanish, ITRW on Disfluency in Spontaneous Speech (DiSS'01), Edinburgh, August, 2001.
Silverman, K., Beckman, M., Pitrelli, Jo., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J. (1992): TOBI: a standard for labeling English prosody. In *ICSLP-1992*, 867-870.