Computational Prosody Starter Bibliography

Pitch Perception . . . Szczepek Reed (2010) chapter 2; Terken (1993); Kochanski (2010); Mertens (2004); Slaney, Shriberg, and Huang (2013); Barnes et al. (2012); Albert, Cangemi, and Grice (2018); Bruggeman et al. (2017)

Voicing Properties . . . Ogden (2017) chapter 4; Esling et al. (2019)

Basic Signal Processing . . . Jurafsky and Martin (in press) section 26.2

Linguistic Prosody . . . Hyman (2006); Wells (2006); Cole (2015)

The Almost-Independence of Prosody . . . Bullyko and Ostendorf (2001); Calhoun and Schweitzer (2012); Torreira and Grice (2018); Vigario, Cruz, and Frota (2019)

Speech Synthesis . . . Zen, Tokuda, and Black (2009); Skerry-Ryan et al. (2018); Lee and Kim (2019); Hodari et al. (2021); Karlapati et al. (2021); Tan et al. (2021)

Speech Recognition . . . Noth et al. (2000); Ostendorf, Shafran, and Bates (2003); Shriberg and Stolcke (2004); Chen et al. (2005); Ward, Vega, and Baumann (2011); Toyama, Saito, and Minematsu (2017); Ryant et al. (2014); Rosenberg (2018)

Paralinguistic Prosody . . . Schuller (2011); Shor et al. (2020)

Historical Context . . . Crystal (1969); Fox (2000)

Prosodic Feature Sets . . . Eyben et al. (2016) sections 3.1 and 3.2; Ward (2019) chapter 8; Kajarekar et al. (2004); Huang, Chen, and Harper (2006); Ferrer, Scheffer, and Shriberg (2010); Eyben, Wöllmer, and Schuller (2010); Degottex et al. (2014); Black et al. (2015); Eyben et al. (2016); Ward (2017); Lenain et al. (2020)

Normalization . . . Sönmez et al. (1997); Shriberg et al. (2000); Ostendorf, Shafran, and Bates (2003); Mariiooryad and Busso (2014)


Turn Taking . . . Skantze (2017); Corps, Gambi, and Pickering (2018); Heldner et al. (2019); Ward and Abu (2016); Ward (2019, 2020); Skantze (2021)

Multistream Temporal Configurations . . . Byrd and Saltzman (1998); Kochanski (2006); Ogden (2010, 2012); Day-O’Connell (2013); Niebuhr (2015); Torreira and Valtersson (2015); Ward (2019); Niebuhr and Neitsch (2019); Niebuhr (2019)

Prosodic Meanings and Functions . . . Hedberg et al. (2010); Kurumada, Brown, and Tannenhaus (2012); Kurumada and Clark (2017); Grice et al. (2017); Ward (2019); Ward and Jodoin (2019)

Superpositional Modeling . . . Kochanski and Shih (2003); Fujisaki (2004); van Santen, Mishra, and Klabbers (2004); Gubian, Boves, and Cangemi (2011); Xu (2011); Liu and Xu (2016); Gerazov and Bailly (2018); Ward (2019)


Individual Variation . . . Pierrehumbert and Steele (1989); Shriberg et al. (2001); Weber et al. (2002); Shriberg et al. (2005); Yoon (2010); Barnes et al. (2012); Niebuhr et al. (2011); Bruggeman et al. (2017); Jun and Bishop (2015); Cangemi, Krüger, and Grice (2015); Roy, Cole, and Mahrt (2017); Boll-Avetisyan, Bhatara, and Höhle (2017); Roessig, Mucke, and Grice (2019); Kim (2019); Xie, Buxó-Lugo, and Kurumada (2021); Wong et al. (2020)

Cognitive and Behavioral Modeling . . . Wagner and Watson (2010); Xu (2015); Byrd and Kri-vokapić (2021)

Unsupervised and Self-Supervised Methods . . . Greenberg et al. (2009); Gubian, Cangemi, and Boves (2010); Neiberg, Salvi, and Gustafson (2013); Janssoone et al. (2016); Madaio et al. (2017); Obin and Beliao (2018); Ward (2019); Wang et al. (2018); Shor et al. (2020)

Challenges . . . Mennen (2015); Ward and DeVault (2016); Niebuhr and Ward (2018); Rosenberg (2018); Marge et al. (2022)

September 9, 2022


Boll-Avetisyan, N.; Bhatara, A.; and Hohle, B. 2017. Effects of musicality on the perception of rhythmic structure in speech. Laboratory Phonology 8(1).


Greenberg, Y.; Shibuya, N.; Tsuzaki, M.; Kato, H.; and Sagisaka, Y. 2009. Analysis on paralinguistic prosody control in perceptual impression space using multiple


Skantze, G. 2021. Turn-taking in conversational sys-


Yoon, T.-J. 2010. Speaker consistency in the realization of prosodic prominence in the Boston University radio speech corpus. In *Speech Prosody, Fifth International Conference*.