

Prosody: Models, Methods and Applications

Nigel G. Ward

University of Texas at El Paso

nigelward@acm.org

Gina-Anne Levow

University of Washington

levow@uw.edu

1 Motivation

Prosody is essential in human interaction, enabling people to show interest, establish rapport, efficiently convey nuances of attitude or intent, and so on.

This tutorial will overview the computational modeling of prosody, including recent advances and diverse actual and potential applications.

We define prosody broadly, as the aspects of spoken utterances that are not governed by segmental contrasts. Some applications that exploit prosodic knowledge have recently shown super-human performance, and our ability to effectively model prosody is rapidly advancing. Yet prosody remains challenging to work with because it operates close to the limits of conscious introspection, and because most spoken utterances involve multiple prosodic dimensions simultaneously serving multiple communicative functions. Intuitions about prosody are often a weak guide for applied work, but a little bit of basic knowledge can go a long way.

2 Outline

Human Fundamentals: articulatory and perceptual aspects [20 minutes]

Processing Fundamentals: prosodic feature computation and feature sets, including issues of individual variation and normalization [30 minutes]

Phonological and Structural Aspects: tone, stress, boundaries, etc. [40 minutes]

Paralinguistic Functions [20 minutes]

Pragmatic Functions, including turn taking, topic structuring, and stance taking functions [70 minutes]

... and interleaved with the above ...

Representations, Models, and Algorithms,

including such recent developments as superpositional modeling, the use of unsupervised methods, and sequence-to-sequence algorithms

Current Trends, including modeling prosody beyond just intonation, representing prosodic knowledge with constructions of multiple prosodic features in specific temporal configurations, and modeling multispeaker phenomenon

Historical Perspectives, briefly, including the long view but focusing on the last 5-10 years

Tools and Resources, and common pitfalls in their uses

Challenges, both short term and long term

Applications, including speech synthesis, speech recognition, diagnosis of medical conditions, inference of speaker sentiments, states and intentions, adaptation in dialog, information retrieval, speaker identification, skills training and assessment

Short Exercises (non-computational)

Throughout, diversity will be a recurring theme, in terms of the different ways in which prosody serves different kinds of functions, in terms of differences in prosodic behaviors across genres, in terms of prosody in typologically-different languages, and in terms of diverse applications.

3 Target Audiences

We envisage three main audiences.

1. Many students of computational linguistics have little exposure to prosody, and what they do

learn is usually 10 to 20 years out of date. There are great opportunities in industry for speech scientists and engineers (as distinct from language scientists and engineers in general) with unmet needs in the tech giants, in traditional industries, and in start ups. The rise of conversational agents has greatly increased student interest in speech, and we hope that our tutorial will help satisfy their curiosity and open doors for students who might not otherwise even be considered for positions in this field. While today most aspects of speech processing are handled by algorithms which are also used for other computational linguistics purposes, prosody, as a phenomenon entirely unique to the spoken language, has different properties and different functions from the rest of language, and is thus possibly the most important aspect of speech for students to learn about.

2. Developers of language processing applications can easily over- or under-estimate the power of prosody and the ease of using it. In this tutorial we will aim to give participants the ability to, given an application potentially exploiting prosody, evaluate the relevance, feasibility and likely value of various approaches and methods.

3. Research team leaders and Ph.D. students may consider starting a research project that involves prosody, whether centrally or marginally. This tutorial will identify key opportunities, issues, and challenges.

But almost anyone in computational linguistics may benefit from this tutorial, as prosody is a topic of wide cross-cutting relevance, including to grammar, discourse, pragmatics, nonverbal communication, and language learning. Considering the roles and nature of prosody may provide insight and new ways to look at both classic problems and emerging applications, such as those involving multimodalilty, hard realtime performance, and perceptions of systems as human-like agents.

This tutorial will be at an introductory level, assuming no previous knowledge of prosody. We expect that most participants will be familiar with basic issues in modeling language and in standard methods for learning from data, but no specific knowledge will be assumed. Familiarity with basic phonetics and phonology would be helpful, but is again not assumed.

4 Small Reading List

An Introduction to English Phonetics, 2nd Edition, Richard A. Ogden, Edinburgh University Press, 2017. Chapter 4.

Analysing Conversation: An introduction to prosody. Beatrice Szczepek Reed. Palgrave Macmillan, 2010. Chapter 2.

The Geneva Minimalistic Acoustic Parameter Set. Florian Eyben, Klaus Scherrer et al. *IEEE Transactions on Affective Computing* 7:2, pp 193-194, 2016. sections 3.1 and 3.2.

Prosody in Context: A review. Jennifer Cole. *Language, Cognition and Neuroscience*, 30:1-2, 1-31, 2015.

Speech Prosody: theories, models and analysis. Yi Xu. In *Courses on Speech Prosody*, A. R. Meireles (ed), Cambridge Scholars Press, 2015 pp 146-177.

Prosodic Patterns in English Conversation. Nigel G. Ward. Cambridge University Press, 2019. Chapter 8.

Speech and Language Processing, 3rd Edition draft. Dan Jurafsky and James H. Martin, 2021. Section 26.2

5 Presenters

Nigel Ward
Computer Science
University of Texas at El Paso
nigelward@acm.org
<http://www.cs.utep.edu/nigel/>

Gina-Anne Levow
Linguistics
University of Washington
levow@uw.edu
<http://faculty.washington.edu/levow>

Ward's research interests lie at the intersection of spoken dialog and prosody. His expertise includes applications of prosody in information retrieval, speech recognition, dialog systems, and language learning. He is known for the creation of a robust prosodic feature set for processing prosody in dialog data, for the computational modeling of prosodic constructions, and for data-backed descriptions of the prosody of dialog in English, Mandarin, Spanish and Japanese. He is the author of *Prosodic Patterns in English Conversation* (Cambridge University Press, 2019) and is for 2018-2022 Chair of the Speech Prosody Special

Interest Group of the International Speech Communication Association.

Levov's research concentrates on the use of intonation in spoken dialog, and her interests range over natural language processing, spoken language systems, and human-computer interfaces. Her expertise includes examination of the prosodic correlates of stance taking, modeling dysarthria, describing and modeling endangered languages, identifying the prosodic markers of turn taking in Arabic, Spanish and English, and developing minimally supervised machine learning techniques to recognize lexical tones in Mandarin, Cantonese, isiZulu, and isiXhosa.

6 Resources

Available at <http://www.cs.utep.edu/nigel/intro-to-prosody/> .