



Prosodic Patterns in Dialog

8th ISCA Speech Synthesis Workshop, September 1, 2013

Nigel G. Ward
University of Texas at El Paso
nigelward@acm.org

In human-human dialog, over 80% of the variance in prosody can be explained by just 20 prosodic patterns, most of which involve actions of both speakers and most of which last several seconds. In dialog these patterns frequently occur simultaneously, at varying offsets, and they are additive at the signal level and apparently compositional at the semantic/pragmatic level. These patterns provide a simple, non-structural way to model the prosodic implications of various functions important in dialog, including managing turn-taking, framing topic structure, grounding, expressing attitude, and conveying instantaneous cognitive state. These patterns have been used successfully for language modeling, for detecting important moments in the speech stream, and for information retrieval from audio archives, and may be useful for speech synthesis for dialog applications.

The lists of dimensions below are compiled from three papers — A Bottom-Up Exploration of the Dimensions of Dialog State in Spoken Interaction, Nigel G. Ward and Alejandro Vega, Sigdial 2012; “Towards Empirical Dialog-State Modeling and its Use in Language Modeling”, Nigel G. Ward and Alejandro Vega, Interspeech 2012; and “Where in Dialog Space does *Uh-huh* Occur?” Nigel G. Ward, David G. Novick, Alejandro Vega, Interdisciplinary Workshop on Feedback Behaviors in Dialog, 2012 — plus unpublished work with Alejandro Vega and Luis F. Ramirez.

Summaries of interpretations of some of the top dimensions found in the Switchboard corpus, with the variance explained by each.

1	this speaker talking vs. other speaker talking	32%
2	neither speaking vs. both speaking	9%
3	topic closing vs. topic continuation	8%
4	grounding vs. grounded	6%
5	turn grab vs. turn yield	3%
6	seeking empathy vs. upgraded assessment	3%
7	floor conflict vs. floor sharing	3%
8	dragging out a turn vs. ending confidently and crisply	3%
9	topic exhaustion vs. topic interest	2%
10	lexical access or memory retrieval vs. disengaging	2%
11	low content and low confidence vs. quickness	1%
12	claiming the floor vs. releasing the floor	1%
13	starting a contrasting statement vs. starting a restatement	1%
14	rambling vs. placing emphasis	1%
15	speaking before ready vs. presenting held-back information	1%
16	humorous vs. regrettable	1%
17	new perspective vs. elaborating current feeling	1%
18	seeking sympathy vs. expressing sympathy	1%
19	solicitous vs. controlling	1%
20	calm emphasis vs. provocativeness	1%
21	mitigating a potential face threat vs. agreeing, with humor	< 1%
22	personal stories/opinions vs. impersonal explanatory talk	< 1%
23	closing out a topic vs. starting or renewing a topic	< 1%
24	agreeing and preparing to move on vs. jointly focusing	< 1%
25	personal experience vs. second-hand opinion	< 1%
26	signaling interestingness vs. downplaying the current information	< 1%
29	no emphasis vs. lexical stress	< 1%
30	saying something predictable vs. pre-starting a new tack	< 1%
37	mid-utterance words vs. sing-song adjacency-pair start	< 1%
62	explaining/excusing oneself vs. blaming someone/something	< 1%
72	speaking awkwardly vs. speaking with a nicely cadenced delivery	< 1%

Summaries of interpretations of the top 7 dimensions in the Maptask corpus, with the last column noting correspondences to Switchboard-corpus dimensions.

1	this speaker talking vs. other speaker talking	s1
2	low activity, low rapport vs. highly engaged	new
3	neither speaking vs. both speaking	s2
4	grounding vs. grounded	s4
5	turn grab vs. turn yield	s5
6	topic continuation vs. topic change	s3
7	spatial relationship vs. path focus	new
8	meta-level vs. on-task	new
9	comfortable vs. awkward	new