Automatic Discovery of Simply-Composable Prosodic Elements

Nigel G. Ward

Department of Computer Science, University of Texas at El Paso
nigelward@acm.org

Abstract

As a way to discover the elements of prosody, Principal Component Analysis was applied to several dozen contextual prosodic features, sampled at 600,000 timepoints in dialog data. The resulting components are interpretable as prosodic patterns, including some which involve behaviors of both interlocutors. Examining contexts and co-occurring words, many of these have clear interpretations. This suggests that English has at least several dozen prosodic patterns, each with its own communicative function.

Index Terms: principal components analysis, prosodic elements, prosodic patterns, factors, dimensions, contours, superposition, intonation, modeling, dialog, interaction, pragmatics

1. Introduction

To understand a complex machine, one needs to identify the pieces and how they work together. Classical approaches to prosody have found many likely pieces, including targets, contours, impulses, and events, and much has been written about each. However the question of how these elements are combined has received less attention; many models of prosody are vague here. This is a problem because theories that rely on unexplained mechanisms have little predictive power: they are impossible to test rigorously and potentially falsify [1, 2].

Well-specified descriptions of how prosodic elements combine do exist, for example [3, 4, 5, 6]. However so far these have been worked out only for carefully circumscribed sets of phenomena, in datasets where every other prosody-related factor is controlled. The more general trend is, it seems, to give up on modeling prosody in terms of composable elements, instead using machine learning techniques that operate directly over raw features [7, 8]. While this can be of great practical value, the resulting models are tailored to single applications, are difficult to interpret, and do not much advance our understanding of prosody.

This paper presents a way to outflank the problem of combination mechanisms. The novel idea is to start with a composition rule, and to then use it to infer the elements; the reverse of the classical strategy. This guarantees that these elements will compose simply, with no slack in the model. Originally developed for purely practical reasons [9, 10], this method is here presented as it relates to other approaches, with new visualizations, and with more dicussion of the broader implications and prospects.

2. Principal Component Analysis for Prosody

The fundamental assumption of the method is that superposition is the main combining principle for prosodic elements. Thus the elements, whatever they are, are required to be summable, and, when summed in various combinations and weightings, to fully explain the observed reality. The discovery task is accordingly to infer the underlying elements from data. This is an underconstrained problem; however, the desire for models that minimize the number of elements and maximize their explanatory power leads us to Principal Components Analysis (PCA).

PCA can be described in several ways, but it is helpful to view it as an iterative analysis process. In each stage, PCA finds the factor that explains as much as possible of the observed variation, across many datapoints and many variables. It then subtracts out what that factor explains, finds another factor to explain much of the remaining variation, and iterates. For example, if we are interested in statistics on people, including income, wealth, family size, number of cars, age, education level, food budget, and so on, the first underlying factor may be something like socioeconomic status, the second may be related to age, the third may be gender, and so on. The observed variable values for any datapoint (person) are modeled as linear combinations of the factors, and conversely, one can go from the observed values for any datapoint to the values of the underlying factors trivially, with a simple matrix multiplication.

PCA is good for dealing with variables which are highly correlated and thus mutually partially redundant. This is commonly the situation in prosody, and PCA has indeed been used here, for identifying the prosodic and other vocal parameters relevant to emotional dimensions [11] and to levels of vocal effort [12], for categorizing glottal-flow waveforms [13], for finding the factors involved in boundaries and accents [14], for characterizing ambiance [15], and for purely practical purposes [16, 17, 18, 19]. A related method, Functional Data Analysis (FDA), has also been applied to prosody, including for identification of the key dimensions of variation in pitch contours [20, 21, 22, 23]. Despite all these precursors, our strategy, of using PCA as a way to discover prosodic elements, is something new.

3. Base Features

In our approach, the datapoints input to the PCA are points in time, and the variables are various prosodic features. PCA is then applied to discover the underlying factors, and these are the elements of prosody.

Because a prosodic value at a point in time is meaningless without context, for each datapoint we use several dozen base features to broadly represent the local prosodic context. For example, in addition to the average pitch over the past 50 milliseconds, we also use the average pitch over a 50 millisecond window centered 75 ms in the past, and over a 100 ms window centered 150 ms in the past, and so on, for both past and future windows, spanning about 6 seconds centered around the point of interest. Including such features enables the use of PCA for time-series analysis [24, 20, 21]. Unusually, our features are not uniformly spaced, but are denser closer to the point whose context is being considered, as detailed elsewhere [9, 25].

Given that the features are from the local prosodic context, each PCA-derived factor will represent a patterning of prosodic features over that context: a 'phonological entity with a distinct time course' [1]. For example, one factor has a region of speech with a slowly dropping pitch, followed by a region that is quiet and slow in rate, followed by a second region of speech that has an early fast region, but then turns slower and lower. The bottom half of Figure 1 shows a visualization of this.

Mathematically a factor can be present either with a positive weight or a negative weight. However it can be difficult to intuitively understand the contributions of a factor when it has a negative weight for a given datapoint. Accordingly the discussion below will focus on one or the other of the two poles of each factor (dimension), discussing either the pattern characterizing points that are high on the dimension or the one for low points. For example the pattern in Figure 1 is the high side of dimension 6.

For PCA to work, the base features should be continuous-valued, on scales for which summation is meaningful. The features we have used approach this ideal but imperfectly. For loudness we use log energy normalized per track to correct for different recording conditions and different speakers. For pitch height we use percentile in the distribution of pitch seen for that track, thus again normalizing for speaker. For pitch range, we similarly use the number of percentiles between the highest pitch point in the window and the lowest. For windows without voiced frames, the mean pitch values and ranges are used instead. For rate, we use a simple frame-by-frame energy-shimmer measure. These features are from our standard inventory, chosen for utility for modeling turn-taking prosody and for language modeling. Better choices could certainly be made, but fortunately PCA is robust to imperfect and noisy features. Finally, following standard practice, we z-normalize all features before applying PCA.

In contrast to previous uses, here we apply PCA to dialog data, with featuresets accordingly including features of both speakers' prosodic behavior. As a result, the top factors discoved by PCA tend to be those which explain variance not only in the speaker's behavior but also in the interlocutor's behavior. That is, this predisposes the method to find patterns that have interactional significance, with ones that relate purely to one speaker's behavior destined to rank lower.

Further, again in contrast to previous work, we applied PCA to a heterogeneous and large data set. This is because our aim is to see what we can find, rather than to refine some model or answer some question. In particular, most of our work has been with the Switchboard corpus of American English telephone conversations. In each case we built the models using about 600,000 data points, the maximum our computer's memory could handle for PCA. These were obtained by sampling every 10 milliseconds during 16 dialogs, without regard to any notion of sentence, utterance, or turn. Thus each model is built from about 100 minutes of dialog, across multiple speakers, multiple dialog activities, and multiple topics.

Finally, PCA involves an (optional) interpretation step. While some researchers are happy with fully automatic methods, we think that human judgments are a necessary part of scientific inquiry, and in this respect our method is aligned with classical approaches. We differ in choosing to apply the human interpretation after the data-reduction step, rather than up front.

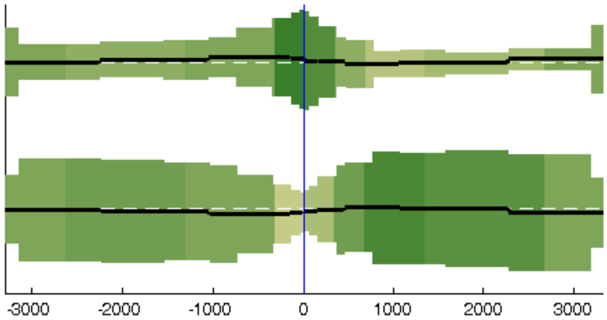

Figure 1: The pattern exemplifying the high side of factor 6 (dimension 6). Time is in milliseconds. First-speaker features are on top and second-speaker features below. Line width shows volume, with the median volume seen at the far left and right of the figure. Height shows pitch, with the median indicated by the dashed line. Darkness shows rate. While what is shown is the strengths of factor loadings — not directly the pitch height, volume or rate — it is not seriously misleading to interpret the figures as indicating typical values for the features across time.

4. Findings

4.1. Interpretable Elements

While PCA is guaranteed to find elements, there is no guarantee that they will be interpretable. While uninterpretable models of prosody can still be useful, interpretable ones are preferable. Luckily, most of the factors that PCA outputs have indeed been interpretable, with each pole corresponding to a simple pattern or 'construction' [26, 27] with an identfiable meaning or function. These generally relate to meanings and functions familiar from the prosody literature. (Although not so far to meanings at the degree of specificity claimed for some sentence-level contours [28, 29, 30].) Space allows just two examples:

4.1.1. The Upgraded-Assessment Pattern

Switchboard dimension 6 is positive to the extent that: the interlocutor was speaking loudly but trails off and this occurs with a low pitch, during which while the speaker was quiet; followed by a loud region by the speaker with a slightly expanded pitch range and increased speaking rate (the upgraded assessment); followed after a short pause by a long and loud continuation by the interlocutor. This is seen in Figure 1.

An example very high on this dimension occurred 309 seconds into dialog sw2402, where A has spoken favorably about warm places:

> A: a lot of people go to Arizona or Florida for the winter and they're able to play all year round but
>
> B: yeah, oh, Arizona's beautiful!

Words frequent in this context include *neat*, *ooh*, *absolutely*, and '*laughter-right*' (laughed tokens of the word *right*).

Thus three kinds of evidence — the feature loadings, impressions of the pragmatics, and statistics of the co-occurring words — provide convergent evidence for a coherent interpretation of the pattern: that it involves one person seeking and the other displaying empathy, in extreme cases in the form of an upgraded assessment. This matches well with a previously-described prosodic construction [31]: a pattern in which a listener expresses agreement with an assessment by producing an upgraded version, for example when one speaker tentatively observes *it's pretty* and the other follows with *absolutely gorgeous* with increased volume, pitch height and pitch range, and 'tighter' articulation. This is exactly what happens when dimension 6 is high.

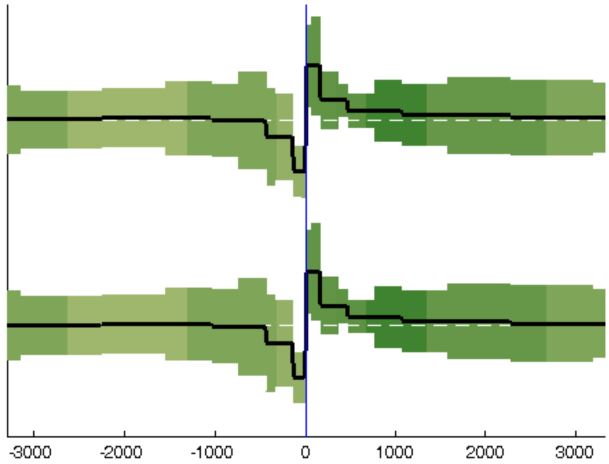

Figure 2: Dimension 26, high side, as above.

4.1.2. *The Interestingness-Signalling Pattern*

Figure 2 shows dimension 26 of the Switchboard corpus. The loadings on this factor do not distinguish between the tracks, but this is just an artifact of the prevalent cross-track bleeding in this corpus, which affects many of the lower dimensions. Examining places in the corpus where this factor was strongly present, we found that these frequently aligned with backchannels. Quantitatively, of all *uh-huh*s in the data, 79% occurred in contexts where factor 26 was present with a positive weight, significantly more than a 50% baseline. Looking at the behavior in the other track, places where factor 26 was high were often places where one speaker was finishing up the delivery of one piece of information and preparing to continue on with some elaboration or new aspect. In general, this pattern of joint behaviors [32] signals that the recent content was interesting and will be interesting again soon.

Looking at the feature loadings, there is a very salient region of low pitch just before the point of interest, for about 150 milliseconds. This is the pattern that previous work has identified as a cue for an interlocutor backchannel [33]. Despite various work aiming to refine and elaborate this cue [34, 35], not much more had been found. However from the figure we can easily read off more information: that this pattern involves a slightly increased pitch for about a second, followed by a short, somewhat louder region (often a content word), followed by a short low-pitch region with reduced volume, and then, about a second after the backchannel, a short region with faster rate (as the speaker resumes the turn with a fresh start). Here PCA serves to reveal the larger pattern encompassing the salient feature.

Just to complete the story, in the opposite pattern, characterizing points where this factor was strongly negative, the speaker was typically involved in a narrative and speaking with low volume, and appeared to be downplaying the importance of what he's saying, for example when it was just background to a main point to come later.

4.2. Numerous Elements

In the quest to reduce prosody to the minimum number of elements, it would be helpful to have estimates of how many elements are really needed. PCA is useful for questions like this: often it reveals that superficially complex phenomena can be explained by just a handful of underlying factors. However that was not the case here; on the contrary, the top 25 factors in Table 1 account for no more than 86% of the variance, despite this featureset being one with many strong correlations. Moreover, at least 30 of the factors (and thus 60 patterns) have clear and distinct functions, as summarized in Table 1. These observations suggest that the research program of reductively explaining prosody [36, 37] may not work for the prosody of dialog.

4.3. Continuous-Valued Elements

A recurring debate in prosody involves the extent to which prosodic elements are categorical or continuous. For these elements we can address this by examining the distributions of values on each dimension. All looked normally distributed, with only two exceptions (on the first dimension, which is bimodal, and on the second, which is skew), which suggests that they are not categorical. This interpretation is compatible with the functions they bear, all of which seem likely to be experienced in a graded rather than categorical manner.

4.4. General Elements

While the elements of prosody are likely to vary somewhat with domain and speaker and so on, it would be disappointing if those found by one application of PCA were entirely limited to one specific genre. To see whether this was the case, we tried it on a different corpus, Maptask, and using a different set of base features (computed using the same feature extractors, but with different densities at different temporal offsets). Again we found meaningful patterns, and of those analyzed so far, most are similar to those found in the Switchboard data, as seen in Table 2.

5. Prospects

Ultimately the aims of prosody research must surely include the identification of the inventory of prosodic elements, for any given language (despite the difficulties [29]). This paper has presented a way to use PCA to advance us towards that goal. Next steps include: 1. using better and finer-grained features, to infer the exact shapes and timings of the patterns. 2. using features that are phrase-, word-or syllable-aligned, rather than fixed in width and offset, 3. analyzing different types of data, to replicate findings obtained with other methods, and 4. examining individual differences, as it is unlikely that all speakers of a language have identical prosodic elements, even if the functions are shared.

In addition to superposition, a complete model will certainly require other combining mechanisms: most obviously concatenation, but also probably stretching, warping, alignment, synchronization, assimilation, undershoot, and others, especially for prosodic phenomena at finer time scales. It important to elucidate how superposition works together with these other mechanisms [1, 40].

Beyond acoustical compositionality, the compositionality of the meanings of these patterns is an open question [41]. The heterogeneity of the functions (Table 1) suggests that they could be composed without mutual interference, but this needs to be

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 [38, 39]	< 1%
62	explaining/excusing oneself *vs.* blaming someone/something	< 1%
72	speaking awkwardly *vs.* speaking with a nicely cadenced delivery	< 1%

Table 1: Brief descriptions of the interpretations of some of the top dimensions found in the Switchboard corpus, with the variance explained by each. Visualizations of all dimensions are at http://www.cs.utep.edu/nigel/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	slowly describing a difficult configuration *vs.* describing an easy path	new
8	meta-level *vs.* on-task	new
9	comfortable *vs.* awkward	new

Table 2: Interpretations of the top dimensions in the Maptask corpus. The last column notes correspondences to Switchboard-corpus dimensions.

investigated.

As noted in the introduction, some recent applications of prosody use raw features directly, without models, or at least without interpretable models. It would be nice to reverse this, to help reunify the scientific study of prosody and practical uses. PCA-derived elements, being computationally convenient yet also interpretable, may help. We already have found them useful for language modeling for speech recognition [10], for information retrieval [42, 25], for finding important information in dialog [43], and for characterizing the pragmatics of a non-lexical discourse particle [44]. Other potential applications include dialog-act inference, simultaneous interpretation, detecting emotion, detecting social roles, language identification, speaker recognition, realtime behavior prediction, dialog outcomes prediction, computer-assisted language learning, language proficiency evaluation, diagnosis of communication disorders, and speech synthesis.

6. Conclusions

Xu and Prom-on envisage models where “a full repertoire of communicative functions can be simultaneously realized in prosody, with all the details of the surface prosody still linked to their proper sources” [6]. PCA-derived prosodic elements can be part of such models, as they meet several important desiderata: 1. a fully explicit composition mechanism for combining elements, here simple addition, 2. groundedness of elements, whose presence at any point in any dataset can unambiguously determined, here by a simple linear combination of easily computable acoustic features, and 3. meaningfulness of elements, here with each bearing a specific communicative meaning or function.

This technique also has other advantages. It works not just for careful, professional speech and the phenomena therein, but for ‘messy’ unconstrained dialog. It covers elements at multiple ‘levels’ with a single mechanism. It’s single mechanism covers not only pitch but also duration and volume (and potentially also voicing modes, gaze, gestures, etc.). Finally, it has been truly useful for discovery, and in this respect the results it gives, and the visualizations they support, are far clearer than those obtained by previous approaches [45, 46, 47, 35]; in essence this is because PCA is good at stripping out the variation involved in dimensions other than the single one being focused on.

Given the simplicity of the method, these results are surprisingly promising and the potential value seems great.

7. Acknowledgments

I thank Olac Fuentes for suggesting PCA, Alejandro Vega for the initial implementation and much more, Luis F. Ramirez for the initial Maptask analysis, Steven D. Werner for analyzing the distributions of values on each dimension, and the NSF for support, as project IIS-0914868.

8. References

[1] J. P. van Santen, T. Mishra, and E. Klabbers, "Estimating phrase curves in the general superpositional intonation model," in *Fifth ISCA Workshop on Speech Synthesis*, 2004, pp. 61–66.

[2] Y. Xu, "Speech prosody: A methodological review," *Journal of Speech Sciences*, vol. 1, pp. 85–115, 2011.

[3] H. Fujisaki, "Information, prosody, and modeling – with emphasis on tonal features of speech," in *Speech Prosody*, 2004.

[4] G. Kochanski and C. Shih, "Prosody modeling with soft templates," *Speech Communication*, vol. 39, pp. 311–352, 2003.

[5] Y. Xu, "Speech melody as articulatorily implemented communicative functions," *Speech Communication*, vol. 46, pp. 220–251, 2005.

[6] Y. Xu and S. Prom-on, "Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model-based stochastic learning," *Speech Communication*, vol. 57, pp. 181–208, 2014.

[7] E. E. Shriberg and A. Stolcke, "Direct modeling of prosody: An overview of applications in automatic speech processing," in *Proceedings of the International Conference on Speech Prosody*, 2004, pp. 575–582.

[8] H. Zen, K. Tokuda, and A. W. Black, "Statistical parametric speech synthesis," *Speech Communication*, vol. 51, pp. 1039–1064, 2009.

[9] N. G. Ward and A. Vega, "A bottom-up exploration of the dimensions of dialog state in spoken interaction," in *13th Annual SIGdial Meeting on Discourse and Dialogue*, 2012.

[10] ——, "Towards empirical dialog-state modeling and its use in language modeling," in *Interspeech*, 2012.

[11] M. Goudbeek and K. Scherer, "Beyond arousal: Valence and potency/control cues in the vocal expression of emotion," *Journal of the Acoustical Society of America*, vol. 128, pp. 1322–1336, 2010.

[12] M. Charfuelan and M. Schröeder, "Investigating the prosody and voice quality of social signals in scenario meetings," in *Proc. Affective Computing and Intelligent Interaction*, 2011.

[13] H. R. Pfitzinger, "Segmental effects on the prosody of voice quality," in *Acoustics'08*, 2008, pp. 3159–3164.

[14] A. Batliner, J. Buckow, R. Huber, V. Warnke, E. Nöth, and H. Niemann, "Boiling down prosody for the classification of boundaries and accents in German and English," in *Eurospeech*, 2001, pp. 2781–2784.

[15] J.-P. Goldman, "Prosodyn: a graphical representation of macroprosody for phonostylistic ambiance change detection," *Proceedings of Speech Prosody*, 2012.

[16] S. Itahashi and K. Tanaka, "A method of classification among japanese dialects." in *EUROSPEECH*, 1993.

[17] Z.-H. Chen, Y.-F. Liao, and Y.-T. Juang, "Prosody modeling and eigen-prosody analysis for robust speaker recognition," *ICASSP*, 2005.

[18] C. M. Lee and S. Narayanan, "Toward detecting emotions in spoken dialogs," *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 293–303, 2005.

[19] D. Jurafsky, R. Ranganath, and D. McFarland, "Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates," *Computer Speech and Language*, vol. 27, pp. 89–115, 2013.

[20] M. Gubian, F. Cangemi, and L. Boves, "Automatic and data driven pitch contour manipulation with functional data analysis," in *Speech Prosody*, 2010.

[21] M. Gubian, L. Boves, and F. Cangemi, "Joint analysis of f0 and speech rate with functional data analysis," in *ICASSP*, 2011, pp. 4972–4975.

[22] B. Parrell, S. Lee, and D. Byrd, "Evaluation of prosodic juncture strength using functional data analysis," *Journal of Phonetics*, vol. 41, no. 6, pp. 442–452, 2013.

[23] O. Jokisch, T. Langenberg, and G. Pinter, "Intonation-based classification of language proficiency using FDA," in *Speech Prosody*, 2014.

[24] I. T. Jolliffe, "Principal component analysis for time series and other non-independent data," in *Principal Component Analysis, 2nd ed.*, 2002, pp. 299–337.

[25] S. D. Werner and N. G. Ward, "Evaluating prosody-based similarity models for information retrieval," in *MediaEval Workshop*, 2013.

[26] J.-M. Marandin, "Contours as constructions," 2006, constructions SV1-10/2006.

[27] N. Sadat-Tehrani, "An intonational construction," *Constructions*, vol. 3, 2008.

[28] M. Liberman and I. Sag, "Prosodic form and discourse function," in *Papers from Tenth Regional Meeting, Chicago Linguistic Society*, 1974, pp. 402–427.

[29] A. Cutler, "The context-dependence of "intonational meanings"," in *Papers from the Thirteenth Regional Meeting, Chicago Linguistic Society*, 1977, pp. 104–115.

[30] N. Hedberg, J. M. Sosa, and L. Fadden, "The intonation of contradictions in American English," in *Prosody and Pragmatics Conference*, 2003.

[31] R. Ogden, "Prosodies in conversation," in *Understanding Prosody: The role of context, function, and communication*, O. Niebuhr, Ed. De Gruyter, 2012, pp. 201–217.

[32] H. H. Clark, *Using Language*. Cambridge University Press, 1996.

[33] N. Ward and W. Tsukahara, "Prosodic features which cue backchannel responses in English and Japanese," *Journal of Pragmatics*, vol. 32, pp. 1177–1207, 2000.

[34] L.-P. Morency, I. de Kok, and J. Gratch, "A probabilistic multimodal approach for predicting listener backchannels," *Autonomous Agents and Multi-Agent Systems*, vol. 20, pp. 70–84, 2010.

[35] N. G. Ward and J. L. McCartney, "Visualizations supporting the discovery of prosodic contours related to turn-taking," in *Interdisciplinary Workshop on Feedback Behaviors in Dialog*, 2012.

[36] J. Pierrehumbert and J. Hirschberg, "The meaning of intonational contourse in the interpretation of discourse," in *Intentions in Communication*, P. R. Cohen, J. L. Morgan, and M. E. Pollack, Eds. MIT Press, 1990, pp. 271–310.

[37] F. Lie, Y. Xu, S. Prom-on, and A. C. L. Yu, "Morpheme-like prosodic functions: Evidence from acoustic analysis and computational modeling," *Journal of Speech Sciences*, vol. 3, pp. 85–140, 2013.

[38] D. R. Ladd Jr., "Stylized intonation," *Language*, pp. 517–540, 1978.

[39] J. Day-O'Connell, "Speech, song, and the minor third: An acoustic study of the stylized interjection," *Music Perception*, vol. 30, no. 5, pp. 441–462, 2013.

[40] S. Tilsen, "A dynamical model of hierarchical selection and coordination in speech planning," *PLOS ONE*, vol. 8, no. 4, 2013.

[41] C. Portes and C. Beyssade, "Is intonational meaning compositional?" *Verbum*, vol. XXXIV, 2014.

[42] N. G. Ward and S. D. Werner, "Using dialog-activity similarity for spoken information retrieval," in *Interspeech*, 2013.

[43] N. G. Ward and K. A. Richart-Ruiz, "Patterns of importance variation in spoken dialog," in *14th SigDial*, 2013.

[44] N. G. Ward, D. G. Novick, and A. Vega, "Where in dialog space does uh-huh occur?" in *Interdisciplinary Workshop on Feedback Behaviors in Dialog, at Interspeech 2012*, 2012.

[45] J. Edlund, M. Heldner, and A. Pelcé, "Prosodic features of very short utterances in dialogue," in *Nordic Prosody – Proceedings of the Xth Conference*, 2009, pp. 56–68.

[46] A. Rosenberg, "Classification of prosodic events using quantized contour modeling," in *HLT-NAACL 2010*, 2010, pp. 721–724.

[47] D. Neiberg, "Visualizing prosodic densities and contours: Forming one from many," *TMH-QPSR (KTH)*, vol. 51, pp. 57–60, 2011.