9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

A Corpus-Based Exploration of the Functions of Disaligned Pitch Peaks in American English Dialog

Nigel G. Ward

University of Texas at El Paso, USA

nigelward@acm.org

Abstract

The exact positioning of pitch peaks often has communicative significance, but the meanings and functions this conveys have never been systematically studied. This paper reports an exploration of one basic aspect of this in one language. The phenomenon is that of "disaligned pitch peaks," that is, peaks which, contrary to the usual tendency, are not aligned with a strong energy peak. The language is American English, as used in naturally-occurring two-person interactions. To find examples, I developed a model to automatically estimate the extent to which a speech signal exhibits a strong pitch peak that is not aligned with an energy peak. Examination of examples revealed many associated pragmatic functions, including suggesting, grounding, agreeing with reservations, implying, and expressing liking, most of which have not been previously noted.

Index Terms: prosody, alignment, pitch peak location, pragmatics, delayed peak, peak delay, late peak, tonal alignment, L-H*, L*-H

1. Motivations

Phenomena of pitch peak alignment and pitch shape have been well-studied, with the phenomenon of delayed pitch peaks in particular having received much attention. Much is known about various aspects, including phonetic properties, ways to represent them phonologically, and correlations with speaking rate, social roles, and dialects [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15]. However the functions of pitch-peak positioning have not previously been systematically investigated.

This paper investigates the meanings and functions associated with one general aspect of pitch peak alignment, namely "disalignment," by which I mean the existence of a significant offset between the pitch peak location and the closest energy peak location. The general tendency in English, as in perhaps all language, is for a pitch peak to align with the intensity peak in a stressed syllable. For example, Barnes *et al.* [7] found, in their control condition, that pitch and energy peaks were on average offset by a mere 16 milliseconds. This precision of alignment is impressive, but far from unique: humans are generally able to precisely synchronize articulatory actions [16]. This is true not only for speaking but also for gesture while speaking: not only do gestures frequently align with stressed syllables, the apex of the gesture even tends to align with the exact pitch peak [17, 18]. Humans also strongly tend to produce actions in alignment [19, 20], even when asked to try not to.

However previous work has shown that, in some conditions, people consistently depart from this general tendency. This suggests that disalignment is not just some sloppiness in behavior; but rather that speakers are doing something special to overcome this tendency: effortfully making the pitch peak *not* align with the energy peak in order to convey or do something.

In dialog, disaligned pitch peaks vary greatly in form. Figures 1 and 2 illustrate. Some of these cases have been studied under more specific names. For example, the peaks on the second syllables of *transfer* and *student* may be annotated as L*-H, as these peaks come roughly a syllable later than the location of the citation-form lexical accent annotated. Others might be annotated as L-H*, and the last example, occurring utterance-finally, could be described as an instance of uptalk. Several look like what might be called scooped rises. All of these examples might also be called late or delayed peaks. This paper will examine the general phenomena of disaligned peaks, without distinguishing among subtypes.

While for some of these subtypes the functions have been well-studied, this is not true for all types. Thus the aim of this paper is to explore the meanings and functions of disaligned peaks, in general, in dialog.

2. An Automatic Estimator of Disalignment

Disaligned pitch peaks are the exception rather than the rule. To find a wide set of samples to examine, it is therefore helpful to have a way to find them automatically. Unfortunately no standard methods are suitable for spontaneous data, since previous approaches rely in one way or another on either the existence of a neutral control condition or hand-labeled landmarks. For example, one common way to identify late peaks requires speech recorded under controlled conditions, with each phrase recorded twice, once with an aligned peak and then with a late peak. This enables measurement of how far a pitch peak is shifted relative to its location in the control. Natural dialog, however, contains no matched productions, so there are no neutral controls. Many previous approaches also rely on measuring deviation from some standard position using timepoints identified by hand, for example the time from a landmark, such as the midpoint of the vowel, to an estimate of the peak's perceived location, such as the tonal center of gravity [7]. Other work has used different landmarks — including energy peaks, voicing onset times and pitch-rise start points — and different correlates of peak location — including pitch turning points and fall locations as anchor points or targets [2, 21, 6, 3, 4, 5].

Thus previous methods are not well suited to corpus-based investigation, so we need a new method. A new method should be robust to the observed diversity of configurations of pitch peaks and energy peaks in spontaneous dialog, as illustrated in Figures 1 and 2. One option considered was to build on methods that estimate the location of pitch peaks relative to segment boundaries [22], however these presuppose a speech recognizer to find the segments, something not available for all languages. Fortunately, disalignments are a robust phenomenon in many situations, not over-sensitive to exact definitions, as seen by the diversity of measures used in previous research, which suggests that a simpler method may be adequate.

10.21437/SpeechProsody.2018-71

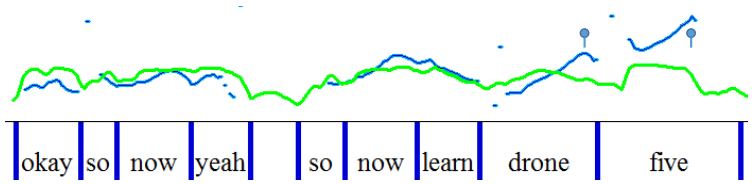

Figure 1: Examples of disaligned peaks, marked with gray balloons. F_0 in blue, energy in green.

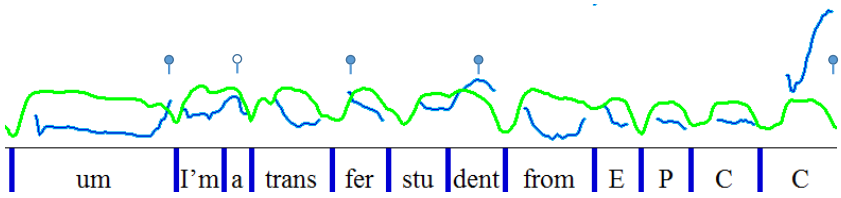

Figure 2: More examples, including a possibly/weakly disaligned peak marked with a white balloon.

Accordingly I built a disalignment estimator that does three simple things: it identifies energy peaks across the data, it independently identifies pitch peaks across the data, and then for each pitch peak it computes a measure of how strongly it avoids aligning with an energy peak. This method is able to produce estimates of the strength of the evidence for a pitch disalignment at any point; I apply it every 10 milliseconds across a corpus. Each of the three things is modeled probabilistically, rather than with discrete decisions, to make the method more robust to noise and microprosodic effects. This method makes no English-specific phonological assumptions, so it should also work for other languages.

For use in spontaneous dialog it is also critical for the method to be robust to pitch doubling, pitch halving, unvoiced regions, and speaker differences in pitch range. These needs are met in part by computing pitch peaks over a representation in which the perceptual value of the height of the pitch at each moment is approximated as the percentile of the F_0 in the speaker's observed F_0 distribution, with unvoiced moments represented as zero. While not a generally-suitable representation, this is advantageous for purposes of identifying peaks.

For something to be a clear pitch peak, it needs to be high in the speaker's range, be in a fairly wide voiced region (rather than being one of a few isolated pitch points), and of course be higher than the pitch in the vicinity. The evidence for a pitch peak at any moment in time is accordingly estimated as proportional to the product of three factors: the pitch height at that point, the amount of voicing in the vicinity (estimated by applying a triangle filter 160 milliseconds wide to a 0/1 representation of whether pitch was detected at each moment), and the "peakiness" as estimated by convolving with a Laplacian of Gaussian (LoG) with σ = 100 ms. Figure 3 shows the shape of this function.

For something to be an energy peak it needs to be loud in the speaker's range, loud relative to nearby syllables, and loud within its own syllable. The evidence is according estimated as the product of three factors: the square root of the normalized log energy at that point, the peakiness at that point as estimated using a LoG filter with σ = 150 ms, and the peakiness as estimated using a LoG with σ = 60 ms.

For there to be a disaligned pitch peak at some time t, there must be a point e nearby where the energy contour exhibits more of a peak that it does at t. Currently the algorithm considers points e within a 120 ms window centered about t. The degree of disalignment at t is estimated as the evidence for a pitch peak at t times the difference between the energy-peak evidence at e and that at t.

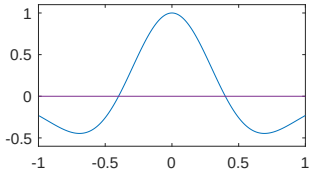

Figure 3: A Laplacian of Gaussian.

The code for this is publicly available [23]. The specific parameters were chosen with reference to a small test set: they resulted in high estimates of disalignment for obvious examples of late peaks, and very low values almost every else. However the behavior was not that sensitive to the exact parameter values. To further gauge the suitability of this method, I applied it to fifteen minutes of varied English and Japanese data, including some examples from [7], and then listened to a sampling of places where it found strong evidence for disaligned peak. About 95% of these looked like late peaks, of various types, as described in the literature. Most exceptions were early pitch peaks: places where a syllable's pitch peak was before the energy peak, most often on ingressive laughter syllables. (While early peaks seem likely to have their own significance [24], my data lacked enough examples to cast light on this.) So the estimator seems reliable enough to use as a tool for locating likely instances of late peak.

Nevertheless it cannot be entirely accurate, for many reasons, including pitch tracker errors, and other factors that cause differences between measured and perceived peak positions, including segmental factors, articulatory constraints, aerodynamic effects, and so on [25, 26, 27]. For this reason, I spotchecked many examples suggested by the model by listening and by examining the pitch and energy contours.

3. Data

I used dialogs from four corpora: two collections of face-to-face conversations [28, 29], one of telephone conversations [30],

and one of people cooperatively playing a two-person video game [31]. The vast majority were in the General American dialect and recorded in Texas, with the telephone conversations recorded mostly in the Dallas area and the other three in El Paso.

In each of the dialogs I used the estimator to compute the evidence for disalignment at every timepoint, and then sorted to find the top few timepoints, excluding those within 2 seconds of a more highly-rated timepoint. These were then sampled for examination.

4. First-Phase Observations

In the first phase of analysis I listened to samples to identify what sorts of meanings and functions were associated with the disaligned peaks, considering the prosody, lexical content, and wider context [32, 33]. This process was not based on any theoretical framework: the only knowledge used was what I know as a naive native speaker of English. I found that many samples were similar to others in meaning or function, so I used qualitative inductive methods to group these into categories [34]. In some cases I inferred a category from only one or two examples, if the function was obvious and seemed to be directly signaled by the late peak.

Below are these general categories, with illustrations for some of the more common ones, chosen to concisely suggest the context and function. The audio for these examples is available[1]. The dark balloons mark the approximate locations of a disaligned peak that made the top 10 for that dialog side, and the transparent balloons other saliently disaligned peaks.

1) Questions, and not only in the final rises, for example in *yeah*↑*?*, in *oh*↑ *is, do you play Halo too*↑*?*, and in *How'd*↑ *your parents get to Ohi*↑*o then?*

2) Grounding, as in Figure 2, where the speaker is introducing into the common ground the referent *EPCC* and the fact that she is a transfer student, and in *the Fire*↑*boy, can touch the fire*, where a player is explaining the videogame characters and their abilities.

3) Partial Agreement, as in the response *yeah*↑ to a question about the recording equipment, to mean "yes, but not really well."

4) Laughing and laughed speech, as in *oh, oh, I thought I was*↑ *going to fall*, and in *I think I'll cut*↑ *that*↑ *part*↑ *out*.

5) Suggestions, as in Figure 1, in *I*↑ *learn half the stuff that I do from just messing around on YouTube* and in *why*↑ *would you do*↑*five, if you still have one left*↑*?*, with the implication that four courses would be a better choice.

6) Liking, as in *I think she dresses so cute*↑, in *I like*↑ *her style* and in *Well*↑*, honest*↑*ly, when I had my last job, with tu*↑ *|toring, I loved*↑ *teaching; I*↑ *love helping kids.*

7) Implication, as in *it's still*↑ *a class*, implying that the interlocutor should retract his statement that a one-credit class doesn't count.

8) Speculation, as in *if I fin*↑*ish my*↑ *classes*↑ *this sum*↑*mer, I feel wrong*↑ *graduating in May*, and in the game corpus *oh, yeah, I was won*↑*dering if, if I would like kill you or something*.

9) Dispreferred Response, as in *I dun*↑*no, I per*↑*sonally feel that, uh, uh*.

10) Mixed Feelings, as in *I don't know*↑ *if I want to work for a little bit, and then*↑*go back for my Masters*.

Some less common categories were: 11) Hopes and Wishes, 12) Correction of a Misconception, 13) Starting a List, 14) Seeking Confirmation, 15) Seeking Feedback, 16) Reported Speech, 17) Offering or Inviting, and 18) Requesting.

Although it is convenient to list these functions as separate categories, in fact they shade off into each other. Some examples are ambiguous between two; others exhibit two or more simultaneously. For example, back in Figure 1, not only was the speaker grounding *Drone Five*, a proper name, but also making a suggestion. In Figure 2 the speaker was suggesting a new topic, grounding a possibly unfamiliar referent, and seeking feedback. The example of partial agreement with *yeah* also involved correction of a misconception and seeking feedback.

Also common are utterances containing multiple cases of disalignment, a phenomenon not previously noted. These all seem to involve intensified pragmatic force or multiple related functions simultaneously present. This suggests that late pitch peak might be modeled as a gesture or feature with temporal extent, rather than as strictly associated with just one syllable.

5. Second-Phase Observations

To examine a larger set of samples, and to do so more systematically, in the second phase I brought in a linguistically-naive research assistant. She was given 300 timepoints, 10 each in the left and right tracks of 15 natural, unscripted dialogs. For each speaker, these were the 10 places where, according to the model, the peak disalignment was greatest. Thus they included very few of the weaker examples, such as the first one in Example 2, for the conversation corpora. However in the gaming corpus there were fewer strongly-disaligned places, so samples there were mostly weaker. Overall these points represented 30 speakers and about 90 minutes of dialogs. This data was evenly distributed across the three dialog types, from three corpora [28, 30, 31]. There was some overlap between these samples and those that I had examined earlier in the process of developing the categories.

The assistant's task was, for each timepoint, to use Elan to navigate to that timepoint in the audio, identify the enclosing utterance, write down the words, and "Listen to the context to discover the dialog activity or discourse role of that utterance, that is, what the speaker is doing with those words." She was further instructed: "If the activity or activities you noticed are on the list of Discourse Functions [the 18 listed above], write down those that are present .... If the utterance is doing something not on the list, or if it is additionally doing something not on the list, write a brief description of what that is." The list she was given consisted of Functions 1 through 18 above, each explained with a phrase or two.

The primary result was confirmation of the relevance of these categories. As seen in Table 1, the majority of these samples were associated with one or more of the 18 functions: 199 out of 300 examples. Because some utterances involved multiple functions, the column totals in Table 1 exceed the number of samples. It is important to note that the exact counts of occurrences for each function not significant, as they reflect in some cases the behavior of a single speaker and in other cases the by-chance lack of any dialog activities of a certain type in this subset of the data.

However not all of the samples matched a previously-identified category. There were several common reasons for this.

First, there were many cases where it was entirely unclear what functions, if any, were being served by an utterance. This was an issue especially in the gaming corpus, where the dialog was sporadic and mostly incidental to the gameplay.

[1] http://www.cs.utep.edu/nigel/disaligned/

function	occurrences f2f	tel.	game	total
question	11	15	12	38
grounding	21	9	4	34
partial agreement	10	17	1	29
laughter	14	4	11	29
suggestion	2	6	17	25
approval, liking	4	6	6	16
implication	6	0	6	12
speculation	5	5	1	11
disprefered response	0	9	0	9
mixed feelings	3	6	0	9
correction	4	1	1	6
hopes or wishes	3	1	1	5
seeking confirmation	2	2	0	4
starting a list	1	2	0	3
seeking feedback	1	0	0	1
reported speech	1	0	0	1
offering or inviting	0	0	0	0
requesting	0	0	0	0
distress	0	0	14	14
telling a story	8	0	0	8
none of the above	16	28	35	79
total	112	112	109	333

Table 1: Counts of occurrences of some functions observed in the three corpora. f2f = face to face; tel. = telephone

Second, there were a few instances of sarcasm, for example, *I love↓ ethics*, where the prosody was used to convey a meaning opposite to its usual meaning.

Third, there were a few instances that were not actually late peaks.

Fourth, by examining the remaining exceptions, we identified two new categories:

19) Distress, as in *Oh, no!↓ I'm so↑ scared*, in the video game corpus,

20) Telling a Story, as in *well↓ okay, so I guess that's a story* and in *I think this tree↓ was probably twice that size* and in *I'll tell↓ you this right now: air↑bags hurt.*

With these two new categories, of the 300 places examined, 221 (74%) exhibited at least one of these 20 functions.

6. Summary and Open Questions

Overall, as a result of this, the first systematic study of the functions of disaligned pitch peaks in dialog, I conclude that the functions associated with late pitch peak in English seem to be much broader than previously thought. It is likely that additional functions remain to be identified, such as perhaps apologies and declining offers [35]. Many directions for follow-on work beckon.

One direction for future work would be to investigate the perceptual space, since clearly that not all disaligned peaks are perceived the same way or as strongly or have the same significance [36]. One special topic in this area would be to investigate the perceptions of late peak as a component of laughter, given that we already know that many other phonetic components of laughter convey some meaning or function [37, 38, 39, 40]. Another topic in this area would be, since casual observation suggests that higher, later, and more numerous late peaks in an utterance may convey the associated functions more strongly, to investigate what specific properties of disaligned peaks, if any, correlate with percepts such as degree of suggestiveness or degree of approval. In particular, since the automatic method presented here can model not only the presence but also the strength of disalignment, it may be useful for deepening our understanding of the categorical and gradient aspects of peak alignment perceptions [27, 10, 26]. Finally, work clarifying the percepts associated with disaligned peaks could lead to a set of data to meet the burning need for a standard reference for testing or calibrating detectors for late peak or disaligned peak.

Another direction for future work would be to investigate the semantic field of functions associated with disaligned peaks. The functions identified above are closely related to those described in the literature for uptalk and delayed peaks. For example, incredulity [2] relates to grounding, partial acceptance and questioning; persuasion [10] relates to suggestions and grounding; and both topic start [9] and new information [41] relate to grounding. Further, although not described in these terms, some of the L*+H examples of [42] look like examples of partial agreement, grounding, making suggestions, and inviting implications. It would be interesting to investigate how these all relate, and whether they might be rooted, in some way, in one or more deeper underlying meanings.

A third direction would be to systematically examine the contexts of occurrence of disaligned peaks, both the discourse contexts and the immediate local contexts, as it seems that in some cases they bear meaning not by themselves, but by appearing concatenated with or superimposed on other prosodic forms [42, 35].

Ideally these further investigations should be carried out in concert, so that, for example, we can identify which specific disaligned forms convey which specific meanings in which specific contexts.

While the theoretical interest of peak alignment phenomena has long been recognized, their practical importance has seemed small: they have been neglected by applied research, with perhaps only one exception, in information retrieval [41]. This is understandable: with the best-known meanings rare ones like incredulity, disalignment phenomena have seemed mostly irrelevant, and left out of modeling work and even prosodic feature sets. For example, the popular GeMAPS set, designed for paralinguistic inference from prosody, treats energy and pitch features as separate streams, meaning that it is unable to capture alignment phenomena at all [43]. However, the current study has identified disalignment-related functions that include aspects that are very widely relevant — including for better second language teaching to foster communicative effectiveness, improving speech synthesis to more effectively convey dialog intentions, improving dialog act recognition, and better tracking of user intentions and attitudes in dialog systems [44]. Thus future work aiming to better understand peak alignment phenomenon is likely to have not only theoretical importance but also great practical value.

7. Acknowledgements

This work was begun at Kyoto University. I thank Y. I. Ward for assistance, Alejna Brugos for sharing data, and Anindita Nath for discussion. This work was supported in part by DARPA under the Lorelei program and by a Fulbright Award. This work does not necessarily reflect the position of the Government, and no official endorsement should be inferred.

8. References

[1] D. R. Ladd, "Phonological features of intonational peaks," *Language*, pp. 721–759, 1983.

[2] J. B. Pierrehumbert and S. A. Steele, "Categories of tonal alignment in English," *Phonetica*, vol. 46, no. 4, pp. 181–196, 1989.

[3] Y. Xu, "F_0 peak delay: When, where, and why it occurs," in *International Congress of the Phonetic Sciences*, 1999, pp. 1881–1884.

[4] J. P. H. van Santen, E. Klabbers, and T. Mishra, "Toward measurement of pitch alignment," *Italian Journal of Linguistics*, vol. 18, pp. 161–187, 2006.

[5] O. Niebuhr, M. D'Imperio, B. G. Fivela, and F. Cangemi, "Are there shapers and aligners? Individual differences in signalling pitch accent category," in *International Congress of Phonetic Sciences*, 2011, pp. 120–123.

[6] Y. Hasegawa and K. Hata, "The function of F0-peak delay in Japanese," in *21st Annual Meeting of the Berkeley Linguistics Society*, 1995, pp. 141–151.

[7] J. Barnes, N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel, "Tonal center of gravity: A global approach to tonal implementation in a level-based intonational phonology," *Laboratory Phonology*, vol. 3, pp. 337–382, 2012.

[8] P. Prieto, J. Van Santen, and J. Hirschberg, "Tonal alignment patterns in Spanish," *Journal of Phonetics*, vol. 23, pp. 429–451, 1995.

[9] A. Wichmann, J. House, and T. Rietveld, "Peak displacement and topic structure," in *Intonation: Theory, Models, and Applications*. ISCA, 1997, pp. 329–332.

[10] A. Arvaniti and G. Garding, "Dialectal variation in the rising accents of American English," in *Papers in Laboratory Phonology 9*. Mouton de Gruyter, 2007, pp. 547–576.

[11] P. Prieto, "Tonal alignment," in *The Blackwell Companion to Phonology*, M. Oostendorp *et al.*, Eds. Blackwell-Wiley, 2011, pp. 1185–1203.

[12] B. Post, M. DImperio, , and C. Gussenhoven, "Fine phonetic detail and intonational meaning," in *Proceedings of The 16th International Congress of Phonetic Sciences*, 2007, pp. 191–196.

[13] C. Graham and B. Post, "Second language acquisition of intonation: Peak alignment in American English," *Journal of Phonetics*, vol. 66, pp. 1–14, 2018.

[14] M. Grice, S. Ritter, H. Niemann, and T. B. Roettger, "Integrating the discreteness and continuity of intonational categories," *Journal of Phonetics*, vol. 64, pp. 90–107, 2017.

[15] D. R. Ladd, A. Schepman, L. White, L. M. Quarmby, and R. Stackhouse, "Structural and dialectal effects on pitch peak alignment in two varieties of British English," *Journal of Phonetics*, vol. 37, pp. 145–161, 2009.

[16] Y. Xu, "Syllable as a synchronization mechanism," in *Proceedings of the 8th Tutorial and Research Workshop on Experimental Linguistics*, A. Botinis, Ed. ISCA, 2017, pp. 15–18.

[17] A. Kendon, "Gesticulation and speech: Two aspect of the process of utterance," in *The relationship of verbal and nonverbal communication*, M. R. Key, Ed. de Gruyter, 1980, pp. 207–227.

[18] D. P. Loehr, "Temporal, structural, and pragmatic synchrony between intonation and gesture," *Laboratory Phonology*, vol. 3, pp. 71–89, 2012.

[19] B. Parrell, L. Goldstein, S. Lee, and D. Byrd, "Spatiotemporal coupling between speech and manual motor actions," *Journal of Phonetics*, vol. 42, pp. 1–11, 2014.

[20] J. Krivokapic, "A kinematic study of prosodic struture in articulatory and manual gestures," *Laboratory Phonology*, vol. 8, pp. 1–26, 2017.

[21] K. Silverman and J. Pierrehumbert, "The timing of prenuclear high accents in English," in *Papers in Laboratory Phonology I*. Cambridge University Press, 1990, pp. 72–106.

[22] V. Mitra and E. Shriberg, "Effects of feature type, learning algorithm and speaking style for depression detection from speech," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 4774–4778.

[23] N. G. Ward, "Midlevel prosodic features toolkit," 2017, https://github.com/nigelgward/midlevel.

[24] K. J. Kohler, "The linguistic functions of F0 peaks," in *11th Congress of the Phonetic Sciences*, 1987, pp. 149–152.

[25] Y. Xu, "Fundamental frequency peak delay in Mandarin," *Phonetica*, vol. 58, pp. 26–52, 2001.

[26] G. Kochanski, "Prosodic peak estimation under segmental perturbations," *Journal of the Acoustical Society of America*, vol. 127, pp. 862–873, 2010.

[27] O. Niebuhr and H. Pfitzinger, "On pitch-accent identification: The role of syllable duration and intensity," in *Speech Prosody*, 2010.

[28] N. G. Ward and S. D. Werner, "Data collection for the Similar Segments in Social Speech task," University of Texas at El Paso, Tech. Rep. UTEP-CS-13-58, 2013.

[29] N. G. Ward and P. Gallardo, "A corpus for investigating English-language learners' dialog behaviors," University of Texas at El Paso, Department of Computer Science, Tech. Rep. UTEP-CS-15-33, 2015.

[30] J. J. Godfrey, E. C. Holliman, and J. McDaniel, "Switchboard: Telephone speech corpus for research and development," in *Proceedings of ICASSP*, 1992, pp. 517–520.

[31] N. G. Ward and S. Abu, "Action-coordinating prosody," in *Speech Prosody*, 2016.

[32] A. J. Wootton, "Remarks on the methodology of conversation analysis," in *Conversation: An interdisciplinary perspective*, R. Derek and P. Bull, Eds. Multilingual Matters, 1989, pp. 238–258.

[33] N. G. Ward and P. Gallardo, "Non-native differences in prosodic construction use," *Dialogue and Discourse*, vol. 8, pp. 1–31, 2017.

[34] J. B. Bavelas, "Quantitative versus qualitative," in *Social approaches to communication*, W. Leeds-Hurwitz, Ed. Guilford Press, 1995, pp. 49–62.

[35] N. G. Ward, *The Prosodic Pattterns of English Conversation*. Cambridge University Press, 2018, to appear.

[36] A. Ritchart and A. Arvaniti, "The form and use of uptalk in Southern Californian English," in *Proceedings of Speech Prosody*, 2014, pp. 20–23.

[37] S. Kori, "Perceptual dimensions of laughter and their acoustic correlates," in *XIth International Congress of the Phonetic Sciences*, 1987, pp. vol. 4, 67.4.1–67.4.4.

[38] D. E. Mowrer, L. L. LaPointe, and J. Case, "Analysis of five acoustic correlates of laughter," *Journal of Nonverbal Behavior*, vol. 11, pp. 191–199, 1987.

[39] E. Lasarcyk and J. Trouvain, "Imitating conversational laughter with an articulatory speech synthesizer," in *Interdisciplinary Workshop on The Phonetics of Laughter*, 2008, pp. 43–48.

[40] J. Oh, E. Cho, and M. Slaney, "Characteristic contours of syllable-level units in laughter," in *Interspeech*, 2013.

[41] N. G. Ward, J. C. Carlson, and O. Fuentes, "Inferring stance in news broadcasts from prosodic-feature configurations," *Computer Speech and Language, submitted*, 2017.

[42] 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.

[43] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan *et al.*, "The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing," *IEEE Transactions on Affective Computing*, vol. 7, pp. 190–202, 2016.

[44] N. G. Ward and D. DeVault, "Challenges in building highly-interactive dialog systems," *AI Magazine*, vol. 37, no. 4, pp. 7–18, 2016.