Investigating the role of 'yeah' in stance-dense conversation

Valerie Freeman[1]*, Gina-Anne Levow*[1]*, Richard Wright*[1]*, Mari Ostendorf*[2]

[1]Department of Linguistics
[2]Department of Electrical Engineering
University of Washington
Seattle, WA USA

{valerief,levow,rawright,ostendor}@uw.edu

Abstract

This study investigates characteristics of stance-related discourse function, stance strength, and polarity in uses of the word 'yeah.' In an annotated corpus of 20 talker dyads engaged in collaborative tasks, over 2300 'yeahs' fall into six common stance-act categories. While agreement, usually with weak, positive stance, accounts for about three-quarters of the instances, opinion-offering, convincing, reluctance to accept an idea, backchannels, and no-stance represent other common stance-related uses. We assess combinations of acoustic-prosodic characteristics (duration, intensity, pitch) to identify those which differentiate these stance categories for 'yeah' and to determine how they relate to levels of stance strength and polarity. Differences in vowel duration and intensity help to differentiate these fine-grained functions of 'yeah.' Within the larger agreement category, we can further assess the effects of stance strength and polarity, finding that positive polarity is signaled by higher pitch, lower intensity, and longer vowel duration, while greater stance strength shows higher pitch and intensity. Finally, a small set of negative 'yeahs' is examined for more specific stance functions which may be distinguishable by differing pitch and intensity contours.

Index Terms: Stance-taking, stance acts, conversational speech corpus, acoustic phonetics

1. Introduction

When someone is talking, understanding the full scope of their intended meaning involves more than just comprehending the words and decoding the syntax. It also involves understanding the shades of meaning that are encoded in intonation, prosody, and shifts in pronunciation. One of these layers of meaning is stance-taking: an individual's expression of an attitude (typically positive or negative) toward a particular object, claim, or person [1, 2]. Aspects of pronunciation variation associated with prosody (e.g., vowel duration, speech rate, pitch excursion, and intensity) reliably differentiate levels of stance strength in spontaneous speech at a coarse level of analysis [3, 4, 5]. However, more fine-grained analyses of stance-act categories, and their relationships to acoustic variation in spontaneous conversations, have not yet been undertaken.

Building on previous coarse-grained analysis of stance acts in the ATAROS corpus [4], we identified the cue word 'yeah' as a good candidate for further investigation; it has a high occurrence frequency and is associated with a variety of stance acts in the corpus, ranging from the discourse function backchannel (typically with no stance) to emphatic agreement (strong, positive stance). Cue words like 'yeah,' 'okay,' 'alright,' etc. (sometimes referred to as discourse markers) may convey information about discourse structure or make a semantic contribution [6, 7, 8]. Such multi-function words are particularly useful for exploring prosodic cues to specific layers of meaning. In several studies on cue words, prosodic variation, such as pitch accent type, has been shown to reliably distinguish discourse contributions, such as backchannels, from semantic contributions of affirmative cue words such as 'okay' and 'alright' [7, 8, 9]. In these studies, while lexical context was a good determiner of the role of cue words, acoustic features related to prosody were also well correlated with cue word roles: backchannels typically ended in a rising intonation while agreements and cues to new discourse segments ended in falling intonation; new-segment cues had high intensity while discourse segment closers had very low intensity [8].

Given previous findings on the utility of acoustic-prosodic features in differentiating both stance strength and cue word roles, we propose that one or more such features differentiate stance-related uses of the word 'yeah.' More specifically, we predict that vowel duration, intensity, or pitch patterns are associated with fine-grained differences between stance-act types such as *agreement, opinion-offering, convincing, reluctance to accept an idea,* and *backchannels*. We test this prediction using a large sample of stance-annotated conversations taken from the ATAROS corpus, described in Section 2, and acoustic analyses presented in Section 3. Findings are summarized in Section 4.

2. Corpus and annotation

All measures are taken on the ATAROS corpus, which contains high-quality audio recordings of speaker-pairs (dyads) engaged in collaborative tasks designed to elicit a high density and variety of stance moves [4]. The sample in this study consists of 20 different dyads (7 female-female, 3 male-male, 10 mixed-gender) completing two of the tasks: the Inventory task, in which dyads arrange household items on a map of an imaginary superstore, and the Budget task, in which dyads cut expenses from an imaginary county budget. The interactions are hand-transcribed in Praat [10], and word and phone boundaries are automatically time-aligned to the audio using the Penn Phonetics Lab Forced-Aligner (P2FA [11]). In this sample of 8 total hours of conversation, more than 2650 'yeahs' are uttered.

2.1. Stance annotation

Interactions in the corpus are annotated at two levels. At a coarse utterance level, every "spurt," or stretch of speech between pauses of at least 500 ms, is labeled holistically for stance strength (none, weak, moderate, strong) and polarity (positive,

negative, neutral) [5]. Weighted Cohen's kappas with equidistant penalties are 0.87 for stance strength labels and 0.93 for polarity labels, with the unweighted kappa for combined labels at 0.88. For these annotations to be useful in the current analysis of 'yeah,' which often comprises an intonational phrase attached to an utterance with a separate discourse function, the spurt-level labels are replaced with strength and polarity assessed for each 'yeah' independently.

At a finer-grained level, annotators label only words and phrases which perform 'stance acts' (akin to dialog acts involving stance-taking) in categories such as:

- o Offer of opinion or recommendation (e.g., "I think we should...", "That's really important")
- s Solicitation of opinion or agreement (e.g., "What do you think?" "Is that alright with you?")
- c Convincing/credibility: Support (reasons, evidence, experience) for a stance (e.g., "And that's why...", "I read that...", "I know because I was there")
- a Agreement, acceptance (e.g., "I agree, absolutely")
- d Disagreement, rejection (e.g., "No", "That's not right")
- r Reluctance to accept a stance (e.g., "Well, ... maybe")
- f Hedging or softening of a stance (e.g., "But that's just me", "Well, I don't know, but...")
- t Teamwork/solidarity: Rapport-building, encouragement (e.g., "Good idea", "Now we're getting somewhere!")
- b Backchannels (e.g., "Mm-hm, yeah.")
- 0 No-stance (unlabeled for stance type, e.g., factual questions and answers: "Do you have the paper?" – "Yeah.")

Multiple labels are applied to phrases performing more than one stance act type; e.g., offering a suggestion (o) with questioning intonation to solicit another's opinion about it (s). In the distributions shown in Table 1, about half of the 'yeahs' in categories (o, r, c) are also labeled with type (a); these are not included in the (a) counts since they are indistinguishable from their respective o/r/c categories on all measures (duration, pitch, intensity). Annotators consider both lexical and prosodic information in determining the makeup of stance acts, and each annotation is verified or modified by two additional annotators.

Of all 'yeahs' in the sample, 2475 (93%) fall into the six most common categories (a, 0, b, o, r, c), which have sufficient tokens for further analysis and are used by at least 20 speakers, even after 209 are excluded due to inaccurate forced alignments and other technical problems. As detailed in Table 1, about 75% of 'yeahs' are involved in agreement, as might be expected, while little more than 5% are backchannels. While 'yeah' is a very common backchannel in general, the collaborative tasks in the ATAROS corpus elicit mainly short exchanges rather than the longer turns that encourage backchannels. The proportion seen here is comparable to that found in other collaborative-task-oriented corpora (e.g., the Columbia Games Corpus described in [8]), but lower than that observed for unstructured telephone conversations (e.g., in SWITCHBOARD [12]). The rates are also lower than those observed for the goal-oriented ICSI Meetings [13], but these include both collaborative discussions and reporting-oriented meetings.

3. Analysis

To investigate whether stance type, strength and polarity affect acoustic-prosodic features, we extract speaker-normalized measures of duration, intensity, and pitch for all 'yeah' instances.

Table 1: *Distribution of 'yeahs' by stance type.*

Stance type	N uttered	N analyzed	N speakers
$a-agreement$	1856	1691	40
$0-no\ stance$	264	256	38
$b-backchannel$	139	127	25
$o-opinion$	111	98	32
$r-reluctance$	57	48	26
$c-convincing$	48	46	20
$Totals$	2475	2266	40

Analyses first consider characteristics as a function of type, then strength and polarity, where we control for type using the large *agreement* category except for the rare negative polarity case.

3.1. Stance type

3.1.1. Vowel duration

Vowel duration is compared across tokens via the ratio of the duration of each 'yeah' vowel instance to the mean duration of all vowels for the speaker within the task in which it appears. This normalizes for variations in speech rate between speakers and tasks. Overall, the vowel in 'yeah' is about twice as long as the collective vowel average (duration ratio mean: 2.1). A one-way ANOVA (assuming unequal variance) shows stance type to have a significant effect on vowel duration (F[5,189] = 10.09, $p < 0.001$). Welch's two-sample t-tests reveal two clusters of stance types: *reluctance, agreement, backchannels* (r, a, b) have longer vowel durations (ratio mean: 2.1) which differ as a group from *convincing, opinion, no-stance* (c, o, 0) (ratio mean: 1.8).

3.1.2. Intensity

Intensity is extracted using Praat [10] at every 10ms of word duration and then z-score normalized speaker-internally based on all the speaker's utterances in both tasks. The mean is then calculated over vowel duration. In general, 'yeah' mean vowel intensity is slightly higher than average speaker intensity (mean: 0.37). A one-way ANOVA (assuming unequal variance) shows stance type to have a significant effect on vowel intensity (F[5,191] = 6.59, $p < 0.001$). With stance type categories arranged from highest to lowest intensity: r, c, a, o, b, 0, Welch's two-sample t-tests reveal that *reluctance* (r) differs from all other types (mean 0.70), but the other types (with means ranging from 0.10 to 0.56) differ only from those not immediately adjacent (e.g., *backchannels* (b) differ from all types except its neighbors, *no-stance* and *opinion* (0, o)).

The smoothing-spline ANOVA plot in Figure 1 shows intensity contours of each stance type across word duration. To compare differing word lengths together, the nearest z-score normalized measurement to every decile (10%) of word duration is used, with splines connecting the means at each decile, shown here from 30%-90% of word duration in order to reduce edge effects from the initial glide. The clusters identified by durational differences are indicated by line color: (r, a, b) in black, (c, o, 0) in gray. Congruent with the t-tests for mean intensity, the members of each duration cluster are separated by their intensity contours. In the longer-duration cluster, *reluctance* maintains the highest intensity and shows the most separation from all other types, while *agreement* shows moderately-high intensity and *backchannels* moderately-low. In the shorter-duration cluster, *no-stance* maintains the lowest intensity, while

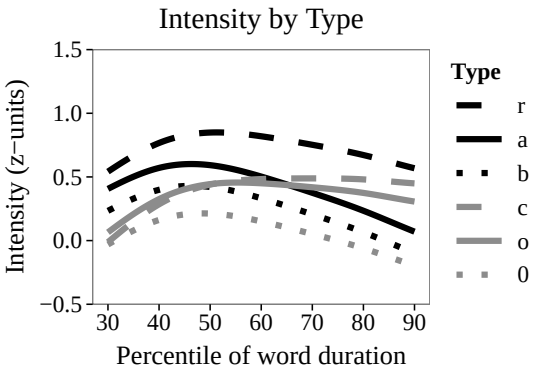

Figure 1: *Intensity by stance type.*

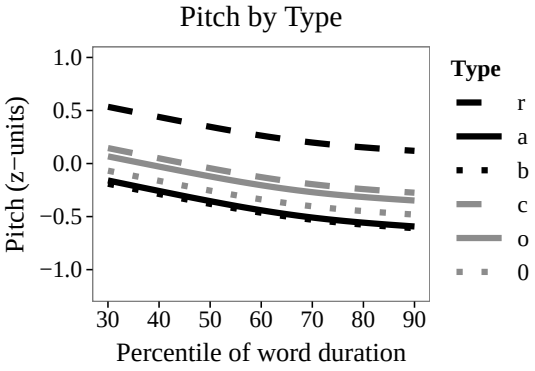

Figure 2: *Pitch by stance type.*

opinion-offering and *convincing* have similar contours which remain flatter after the peak near word midpoint, rather than falling as all other types do. This may be an effect of utterance position, as *opinion-offering* and *convincing* most often appear utterance-initially or -medially, while the other types also end utterances or stand alone as complete utterances.

3.1.3. Pitch

We extract pitch using Kaldi [14][1] at every 10ms of word duration and then log-scale and z-score normalize these values speaker-internally, similarly to the case for intensity. Overall, pitch measures do not add much information, other than to confirm that *reluctant* 'yeahs' behave differently than the other types. A one-way ANOVA (assuming unequal variance) shows stance type to have a significant effect on mean word pitch ($F[5,189] = 8.05$, $p < 0.001$). As with intensity, Welch's two-sample t-tests show that *reluctance* differs from all other types, with the highest mean pitch (mean 0.407), while the other categories overlap with their neighbors (means -0.254 to 0.014), as seen in Figure 2. *Backchannels* and *agreement* have the lowest pitch, and the *backchannels* on average lack the final rise observed in other work (e.g., [8]). In addition, *reluctant* 'yeahs' have higher mean and maximum pitch than words immediately preceding them.

[1]The Kaldi option for long-term mean removal was not used due to biases introduced in regions abutting pauses.

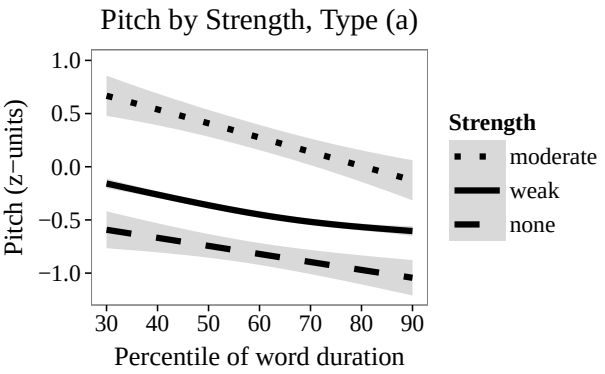

Figure 3: *Pitch by stance strength for agreeing 'yeahs.'*

3.2. Stance strength

Within the category of 1691 *agreeing* 'yeahs,' the majority (1570, 93%) show weak stance strength, with only a few showing no strength (64) or moderate strength (57), and none with strong. Both pitch and intensity separate moderate-strength 'yeahs' from weak and no-strength, which do not reliably differ on aggregate measures. One-way ANOVAs (assuming unequal variance) show stance strength to have a significant effect on mean word pitch ($F[2,79] = 14.14$, $p < 0.001$) and mean vowel intensity ($F[2,84] = 25.65$, $p < 0.001$), but Welch's two-sample t-tests cluster weak and no-strength, separate from moderate. The same pattern holds for pitch minimum, maximum, range, and comparison to immediately preceding words, in which moderate-strength 'yeahs' show slightly higher maximum pitch than their neighbors. Strength levels do not differ by minimum vowel intensity, but maximum intensity increases reliably with each strength level ($F[2,84] = 27.70$, $p < 0.001$).

In addition, all three strength levels show separation throughout their pitch and intensity contours, as seen in the smoothing-spline ANOVA plot in Figure 3, in which gray shading indicates 95% confidence intervals around the mean pitch contours. While all slopes decline over word duration, pitch clearly increases with stance strength. The same scalar relationship holds for intensity (which curves as in Figure 1), although weak and no-strength 'yeahs' show only slim separation.

3.3. Polarity

In the annotation process, speech marked as having stance strength (weak, moderate, strong) is also marked for polarity, i.e., as expressing positive, negative, or neutral sentiment. Unsurprisingly, 'yeah' is usually positive (83% of the analyzed sample), occasionally neutral, showing neither clear positive nor negative stance (16%), and rarely negative (1%). Here we investigate differences between positive and neutral 'yeahs' within the largest stance type category, *agreement*, while the few negative tokens in the sample are examined qualitatively.

3.3.1. Positive vs. neutral

Of 1626 'yeahs' in the *agreement* category with stance strength, 1466 (90%) are positive and 155 neutral. One-way ANOVAs (assuming unequal variance) show that positive 'yeahs' have significantly longer vowel duration ($F[1,183] = 4.03$, $p < 0.05$), pitch ranges that extend significantly higher ($F[1,203] = 18.89$, $p < 0.001$), and a faster intensity drop, which significantly lowers mean vowel intensity ($F[1,191] = 5.31$, $p < 0.05$). The effect

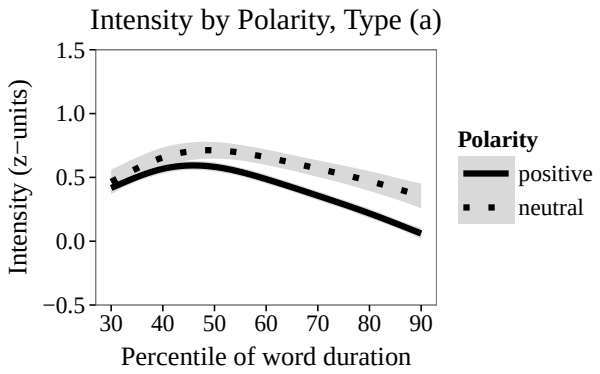

Figure 4: *Intensity by stance polarity for agreeing 'yeahs.'*

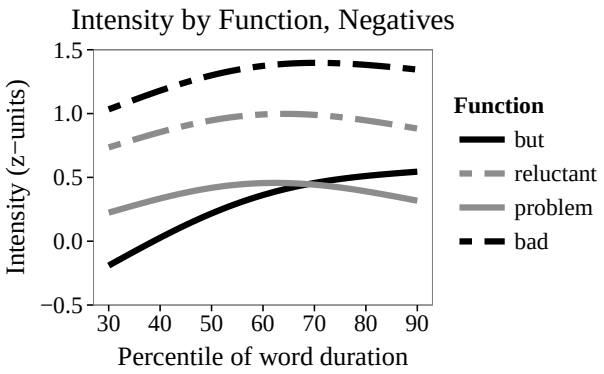

Figure 5: *Intensity by stance function for negative 'yeahs.'*

of intensity can be seen in the smoothing-spline ANOVA plot in Figure 4, in which mean intensity for positive *agreeing* 'yeahs' declines more sharply after word midpoint.

3.3.2. Negative

Negative uses of 'yeah' are rare, so results here are qualitative and must be interpreted with caution, but the findings may guide future work. In a previous stage of analysis on a smaller corpus sample, before the fine-grained stance type annotation had been completed and before strength and polarity were assessed for each 'yeah' independent of its utterance, 43 'yeahs' that occurred in negative utterances were examined for their stance function in a manner similar to later stance type annotation. Four categories of functions emerged from this analysis [15], which were differentiated by their pitch and intensity contours: "tough problem" (an expression of shared difficulty) and "that's bad" (agreement with a negative assessment) group together with lower, flat pitch, while "reluctance" (to accept a stance) and "yeah but" (preceding explanation against a stance) group with higher, curving pitch (dipping and domed, respectively), but "tough problem" and "reluctance" show lower, relatively flat intensity, while "that's bad" and "yeah but" have higher, domed intensity.

In the current, larger sample, after assessing the polarity of each 'yeah' independently of its utterance, only 16 'yeahs' are annotated as expressing negative sentiment. Six of these occur in negative utterances and therefore overlap with the previous data set; the remaining 10 are categorized by stance function according to the scheme applied to the previous sample. This yields 7 *tough problem* 'yeahs,' 4 *yeah but*, 4 *reluctance*, and 1 *that's bad*. While all four in the *reluctance* function category are also annotated for stance type as *reluctance*, the other categories are varied. Each includes *agreement*, *tough problem* includes *reluctance, no-stance*, and *opinion*, and *yeah but* includes *reluctance, no-stance*, and *convincing*. Since components of the two annotation schemes overlap, the mapping between their categories is not one-to-one, but all produce logical pairings, with the possible exception of those marked as *no-stance*. With annotation schemes executed independently, it is plausible that stance type annotation determined that these 'yeahs' did not clearly contribute to a stance, while strength/polarity annotation found them to be weakly negative.

In contrast to the previous sample, the function categories in the current sample are not cross-cut by pitch and intensity contours; rather, they may be divided into two groups: *that's bad* clusters with *reluctance* with both higher pitch and intensity,

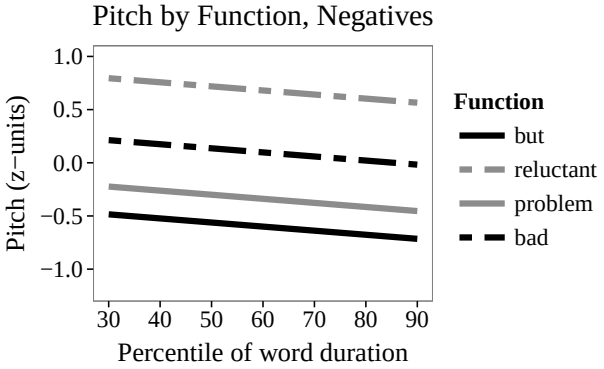

Figure 6: *Pitch by stance function for negative 'yeahs.'*

while *yeah but* and *tough problem* are lower on both measures. In contrast to the domed and dipping contours in the previous sample, all contours in the current sample are fairly level, with pitch declining slightly and intensity rising slightly, with the exception of intensity for *yeah but*, which rises more sharply.

4. Summary

This study investigates acoustic characteristics of stance-related discourse function, stance strength, and polarity in uses of the word 'yeah,' where prosody plays a particularly important role in communicating meaning. In an annotated corpus of over 2300 'yeahs' in dyadic discussions, instances fall into six common stance-act categories: *agreement, opinion-offering, convincing, reluctance to accept an idea, backchannels,* and *no-stance*. The six categories can be distinguished through a combination of intensity contour and duration cues. Pitch is useful for distinguishing strength of stance. Stance polarity is difficult to analyze because of the small number of negative instances, but intensity contour shape and slope appear to be qualitatively different. This categorization and characterization of the functions of 'yeah' lay the groundwork for automatic recognition of the role of this frequent, multi-functional token in conversation.

Acknowledgments

This work is supported by NSF IIS: #1351034. The views/conclusions contained herein are those of the authors and do not necessarily represent the views of NSF or the U.S. Government. Thanks also to our annotators: Heather Morrison, Lauren Fox, Nicole Chartier, Max Carey, and Marina Oganyan.

5. References

[1] D. Biber, S. Johansson, G. Leech, S. Conrad, and E. Finegan, *Longman grammar of spoken and written English*. Longman, 1999.

[2] J. W. Du Bois, "The stance triangle," in *Stancetaking in discourse: Subjectivity, evaluation, interaction*. Amsterdam: John Benjamins Pub, 2007, pp. 139–184.

[3] V. Freeman, "Hyperarticulation as a signal of stance," *Journal of Phonetics*, vol. 45, pp. 1–11, 2014.

[4] V. Freeman, J. Chan, G.-A. Levow, R. Wright, M. Ostendorf, and V. Zayats, "Manipulating stance and involvement using collaborative tasks: An exploratory comparison," in *Proceedings of Interspeech 2014*, 2014.

[5] G.-A. Levow, V. Freeman, A. Hrynkevich, M. Ostendorf, R. Wright, J. Chan, and T. Tran, "Recognition of stance strength and polarity in spontaneous speech," in *Proceedings of the 5th IEEE Workshop on Spoken Language Technology (SLT)*, 2014.

[6] B. J. Grosz and C. L. Sidner, "Attention, intentions, and the structure of discourse," *Computational linguistics*, vol. 12, no. 3, pp. 175–204, 1986.

[7] J. Hirschberg and D. Litman, "Empirical studies on the disambiguation of cue phrases," *Computational linguistics*, vol. 19, no. 3, pp. 501–530, 1993.

[8] A. Gravano, J. Hirschberg, and Š. Beňuš, "Affirmative cue words in task-oriented dialogue," *Computational Linguistics*, vol. 38, no. 1, pp. 1–39, 2012.

[9] A. Gravano, S. Benus, H. Chávez, J. Hirschberg, and L. Wilcox, "On the role of context and prosody in the interpretation of okay," *Proceedings of ACL*, pp. 800–807, 2007.

[10] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [computer program], version 5.3.55," 2013, http://www.praat.org.

[11] J. Yuan and M. Liberman, "Speaker identification on the SCOTUS corpus," in *Proceedings of Acoustics '08*, 2008.

[12] J. Godfrey, E. Holliman, and J. McDaniel, "SWITCHBOARD: Telephone speech corpus for research and development," in *Proceedings of ICASSP-92*, 1992, pp. 517–520.

[13] N. Morgan, D. Baron, J. Edwards, D. Ellis, D. Gelbart, A. Janin, T. Pfau, E. Shriberg, and A. Stolcke, "The meeting project at ICSI," in *Proceedings of Human Language Technologies Conference*, 2001.

[14] P. Ghahremani, B. BabaAli, D. Povey, K. Riedhammer, J. Trmal, and S. Khudanpur, "A pitch extraction algorithm tuned for automatic speech recognition," in *Proceedings of ICASSP 2014*, 2014.

[15] V. Freeman, R. Wright, and G.-A. Levow, "The prosody of negative 'yeah'," in *The LSA Annual Meeting Extended Abstracts (ExtAbs)*, 2015.