JOURNAL ARTICLE

Prosodic features of stances in conversation

Valerie Freeman

Department of Communication Sciences and Disorders, Oklahoma State University, Stillwater, OK, US
valerie.freeman@okstate.edu

Stance—attitudes and opinions about the topic of discussion—has been investigated textually in conversation analysis, discourse analysis, and computational models, but little attention has focused on its acoustic-phonetic properties. It is a challenge, given the complexity of stance and the many other types of meaning that share the same acoustic channels, all overlaid on the lexical and syntactic material of the message. With the goal of identifying automatically-extractable, acoustically-measurable correlates of stance-taking, this work identified signals of stance in prosodic measures of fundamental frequency, intensity, and duration in an audio corpus of dyads engaged in collaborative conversational tasks designed to elicit frequent changes in stance at varying levels of involvement. The study examined over 32,000 stressed vowels in content words spoken by 40 American English-speakers and found that f0 and intensity increased with stance strength, longer vowel duration signaled positive polarity, and a combination of measures distinguished several stance-act types, including: general agreement, weak-positive agreement, rapport-building agreement, reluctance to accept a stance, stance-softening, and backchannels. These results contribute to the understanding of the acoustic-phonetic properties of the social and attitudinal messages conveyed in natural speech.

1. Introduction

Many layers of meaning are conveyed in natural speech, beyond lexical and sentential denotations. One layer is stance, or the expression of an attitude toward an object, claim, or person relevant within the discussion context (Biber, Johansson, Leech, Conrad, & Finegan, 1999; Du Bois, 2007). Stance can be conveyed in many ways, but with only a fraction of the message sent through textual components, much of the information must be present in the delivery, the acoustics of the speech signal itself. Just as changes in pronunciation and prosody can transform a sentence from statement to question, similar changes can affect the intended meaning and reception of social and attitudinal information. Phonetic correlates of information structure, discourse structure, and such social-indexical aspects as the region, gender, ethnicity, or identity of speakers— and perceptions and interpretations of these features by listeners—have been studied in various sociolinguistic and computational fields. However, the phonetic properties of stance-taking have received less attention. This leads to questions of how stance is signaled acoustically. For example, we can express strong or weak opinions, contrast positive and negative attitudes, convey enthusiastic or reluctant agreement, take confident or uncertain positions, engage in persuasion or show deference, all without changing the words we use. How is this accomplished? In addressing this question, this study presents some of the first work to find automatically-extractable acoustically-measurable correlates of stance-taking in natural speech. It employed a large audio corpus of stance-dense collaborative

conversation and identified acoustic-prosodic measures which signal aspects of stance type, strength, and polarity.

1.1. Stance and related work

Stance and related concepts are studied in various disciplines using different terms, including *attitude, evaluation, assessment, appraisal, sentiment,* and *subjectivity* (see Englebretson, 2007 and Jaffe, 2009 for summaries). The work presented here took a broad view of *stance* as used in discourse- and conversation-analytic approaches: "personal feelings, attitudes, value judgments, or assessments" (Biber et al., 1999, p. 966), and of their expression, the social activity of *stance-taking,* also called *evaluation* (Du Bois, 2007; Haddington, 2004; Hunston & Thompson, 2000). Du Bois (2007) described stance-taking as a three-part act which includes evaluation of an object or proposition, positioning of a speaker in relation to that evaluation, and alignment between two speakers and their evaluations. The collaborative tasks analyzed in this study were designed to elicit precisely this process of stance-taking.

Stance-taking is an essential component of interactive collaboration, negotiation, and decision-making. It can involve several levels of linguistic information, including acoustic, prosodic, lexical, and pragmatic elements. Conversation- and discourse-analytic approaches provide many descriptions of stance, often seated in fine-grained content analysis (e.g., Biber & Finegan, 1989; Conrad & Biber, 2000; Du Bois, 2007; Englebretson, 2007; Haddington, 2004; Hunston & Thompson, 2000; Jaffe, 2009). Freeman (2014) drew on such frameworks of stance type classification in order to identify areas of stance-expression for phonetic analysis in an American political talk show. Stance-expressing phrases (e.g., opinions) had faster speaking rates, longer stressed vowels, and more expanded vowel spaces when compared to more neutral phrases (e.g., guest introductions). As a small-scale study, the work examined stance at a coarse level—binary presence/absence, collapsing many categories of stance-taking acts identified in the conversation/discourse-analytic literature. However, different types of stance-taking are likely to have different phonetic correlates, calling for closer inspection.

Conversation analysts have identified prosodic patterns that distinguish stance types within particular contexts. For example, Freese and Maynard (1998) described how opposing uses of prosodic features were associated with deliveries of good versus bad news in conversation. That is, announcements of good news were loud and fast, with high pitch, wide pitch ranges, and frequent pitch rises, while bad news was quiet and slow with low, invariant, and falling pitch contours. Interlocutors' reactions also used these patterns, reflecting the announcer's initial joyful or sorrowful assessments.

Within work on prosody and pragmatics, some studies have considered the prosodic prominence of particular stance-expressing words or phrases. For example, Biber and Staples (2014) found that infrequent stance adverbials in a transcribed corpus of Hong Kong English were used to express attitudes and were normally prosodically prominent, while frequent stance adverbials (e.g., *actually, usually, obviously, maybe, probably*) were used as grammaticalized discourse markers without prominence. Using acoustic software to visualize f0 contours in an audio corpus of British English, Dehé and Wichmann (2010) identified patterns that differentiated uses of *I think* and *I believe* as main clause, comment clause, or discourse marker based on the location of prosodic prominence (pronoun, verb, or neither, respectively). These phrase-level analyses offer valuable insights, but they were limited to specific phrases, and they treated prominence as a confluence of prosodic features without considering the effects of component measures.

Ward, Carlson, and Fuentes (2018) used computational modeling to identify stances in radio news stories, finding that numerous prosodic features interacted to convey

different types of stance. Measures of intensity, pitch height and range, speaking rate, and hyperarticulation were especially useful in locating stances involving assessment (good/bad, praiseworthy/deplorable), newness (new/background, surprising/typical), subjectivity (fact/opinion, controversy), and personal relevance to the audience. Their models considered over 80 different acoustic measures of prosodic features, which has the advantage of testing many complex combinations of features but may have a disadvantage in the human interpretability of the results. The current work took a complementary approach, employing separate acoustic-prosodic measurements (f0, intensity, duration) over all utterances in a conversational corpus. The methods were first employed on a subset of utterances in the corpus, 2266 instances of the word "yeah" (Freeman et al., 2015). In that study, greater stance strength carried higher f0 and intensity; positive polarity was signaled by higher f0, lower intensity, and longer vowel duration; and certain stance types were differentiated by vowel duration and intensity.

The work presented here investigated acoustic correlates of stance-taking in American English conversation with a detailed treatment of stance features, a broad range of stance-expression types, and local phonetic measurements. It took up the argument that since stance presence is signaled acoustically (Freeman, 2014), components or features of stance (strength, polarity, type) are likely to differ acoustically as well (Freeman et al., 2015). The approach leveraged advantages of qualitative content analysis with quantitative phonetic measurement over a sizeable audio corpus of dyadic conversations.

2. Methods

The central prediction of this study was that stance type, strength, and polarity are signaled by changes in the acoustic signal. This prediction was tested using measures of fundamental frequency (f0), intensity, and vowel duration extracted from an eight-hour audio corpus of 40 speakers engaged in collaborative tasks annotated for stance features.

2.1. Data set

The data set for this study was drawn from the ATAROS corpus, a high-quality audio collection of dyads completing collaborative tasks designed to elicit frequent changes in stance (for a full description of the corpus, see Freeman, 2015; for access to the corpus, contact the author). The sample consisted of 20 dyads engaged in two of the tasks, for a total of nearly eight hours of conversation containing over 71,300 words. The acoustic analyses presented here were conducted on lexically stressed vowels within content words (hereafter called 'stressed-content vowels' or SCVs). This was intended to minimize interactions with phonetic reduction typically found in function words and unstressed vowels. SCVs comprised 37% of all vowels in the sample and provided more than 32,000 vowel tokens for analysis.

2.1.1. Speakers

In order to minimize potential stance-related dialect differences, all speakers in the corpus were adult native English-speakers aged 18–75 who grew up in one dialect region, the Pacific Northwest (Washington, Oregon, and Idaho). Ethnicity was not controlled, but the proportions of self-identified ethnicity were consistent with the general ethnic makeup of the Seattle area, where recordings were made (U.S. Census Bureau, 2010). Speakers reported no history of hearing problems, and any speakers with apparent speech impediments were excluded from the current analysis. Dyads were made up of strangers matched roughly by age (within 10 years) and either crossed or matched by sex. **Table 1** shows the distribution of dyads in the sample by age and sex. There were more female

Table 1: Number of dyads by age and sex.

Group	Ages	Dyads by sex		
		FF	MM	MF
Younger	(18–32)	3	1	6
Middle	(38–49)	1	1	3
Older	(60–75)	3	1	1

speakers than male (24 and 16 total), and half were under age 35. Speakers varied in the amount of speech they contributed, but contributions were proportional by sex and age group, with 57% of vowels uttered by females, almost half by the younger group, and about a quarter each for the middle and older groups.

2.1.2. Tasks

Recordings were made in a sound-attenuated booth in a university lab using head-mounted microphones and a separate recording channel for each speaker, resulting in 16-bit stereo WAV-file recordings with a 44.1 kHz sampling rate.

Dyads completed a brief demographic questionnaire and five collaborative problem-solving tasks designed to elicit frequent changes in stance and differing levels of involvement or engagement. The tasks involved two sets of about 50 target items chosen to represent the main vowel categories of Western American English in fairly neutral consonantal contexts (i.e., avoiding liquids and following nasals, which commonly neutralize vowel contrasts; Labov, Ash, & Boberg, 2006). This study analyzed the Inventory and Budget tasks, the two tasks intended to elicit the weakest and strongest stances and levels of involvement, respectively. Both tasks averaged about 13 minutes in duration and about 150 utterances per speaker (for details, see Freeman et al., 2014; Freeman, 2015).

2.1.2.1. The Inventory task

This collaborative decision-making task was designed to elicit low levels of involvement and weak stances. Speakers stood facing a felt-covered wall and were given a box of about 50 Velcro-backed cards that could be stuck to the felt. The cards were printed with the names of household items, and about 15 additional cards were already placed on the wall, which represented a store inventory map. Speakers were told to imagine that they were co-managers of a superstore in charge of arranging new inventory. They discussed each item and decided where to place it on the map. This task generally involved polite solicitation and acceptance of suggestions, as in this example exchange:

A: Books could go near toys I think. Maybe.
B: Yeah or travel guide – Yeah, between toys and travel guides?
A: Yeah, sure.

2.1.2.2. The Budget task

This collaborative decision-making task was designed to elicit high levels of involvement and strong stances. Speakers were seated at a computer screen and told to imagine that they were on a county budget committee in charge of making cuts to about 50 services and expenses. They discussed each item and decided whether to fund or cut it. This task involved more elaborate negotiation, which might include citing personal knowledge or experience as support for stances, as in this excerpt:

A: Well job training programs is pretty crucial. [...] And so is ... chicken pox vaccinations, right?

B: I – well, I didn't get a chicken pox vaccination. I think a lot of kids just naturally get chicken pox and then they're fine.

2.1.3. Annotation

Three levels of manual annotation were conducted: orthographic transcription, stance strength and polarity annotation, and stance type annotation. Annotators were three advanced or recently graduated bachelor's students in linguistics and speech science who were trained and supervised by the author to ensure transcription accuracy and annotation consistency.

2.1.3.1. Transcription

Tasks were manually transcribed in Praat (Boersma & Weenink, 2013) following a simplification of the ICSI Meeting Corpus guidelines (Morgan et al., 2001). Stretches of speech were demarked when surrounded by at least 500 ms of silence, and the resulting 'spurt' was transcribed orthographically using conventional American spelling, with the addition of common shortenings, discourse markers, filled pauses, disfluencies, and vocalizations with clear meanings (Freeman, 2015). Completed manual transcriptions were automatically time-aligned to the audio using the Penn Phonetics Lab Forced Aligner (P2FA; Yuan & Liberman, 2008), which demarked word and phone boundaries for each speaker.

2.1.3.2. Stance strength and polarity

Tasks were manually annotated at a coarse level between pauses for two broad features of stance: strength and polarity. Each spurt (stretch of speech said by one speaker between at least 500 ms of silence) was marked with one of the stance strength labels shown in **Table 2**. Spurts with a discernible stance strength (label 1, 2, or 3) were also labeled for polarity, as shown in **Table 3**. As a result, each spurt was marked with one of 14 possible strength-polarity label combinations.

   Both textual content and prosody were taken into account when determining labels, as prosody can be used to enhance or even reverse the meaning of text alone. One purpose of this study was to identify acoustic cues that people use to convey (and therefore interpret) stance, making it necessary to include the audio signal in the annotation process; however, annotators considered prosody holistically without specific reference to components to be

Table 2: Stance strength levels.

Label	Description and examples
0	Minimal stance: list reading, backchannels, facts (e.g., "Next I have cookies.")
1	Weak stance: cursory agreement, suggesting solutions, soliciting other's opinion, bland opinion/reasoning (e.g., "What do you think?" "Let's do this." "Okay.")
2	Moderate stance: more emphatic versions of items in #1; disagreement, offering alternatives, questioning other's opinion (e.g., "Uh, how about here instead?" "Are you sure?" "Yes! Perfect.")
3	Strong stance: very emphatic versions of items in #1–2 (e.g., "Screw that!" "Oh my god! I can't have that happen on my watch!")
x	Unclear: cannot be determined, excited pronunciations of minimal-stance content (e.g., "Ooh, buckets!" "I don't know what that means.")

measured acoustically (pitch, loudness, duration). Because strength is relative, the scheme was applied on a per-speaker, per-task basis. Before labeling a task, annotators listened to a portion of the task or a prior task to get a general sense of each speaker's styles and strategies. For example, for speakers with small f0 and intensity ranges, small deviations are more meaningful than for more energetic speakers, whose modulations must be more extreme to indicate differences in stance. Annotators listened to both channels of the task audio while labeling one speaker's transcription, and then listened to the task again while labeling the other's.

The scheme was verified for its usability with independent annotation. The first two dyads recorded were used for training and reliability testing. Three annotators independently annotated all four task files with moderately high agreement. Fleiss' kappa was 0.69 for polarity labels, 0.57 for stance strength labels, and 0.55 for combined (strength + polarity) labels. Given the complexity of the annotation task, this level of agreement was deemed sufficient to allow less overlap in annotation in favor of an overall faster procedure. After a task was labeled by one annotator, a second reviewed and verified or corrected each label while listening to the audio and reading the transcript. Asterisks were used to indicate uncertainty, with the second annotator providing a second opinion as needed. If the second annotator remained uncertain about a label, a third annotator served as a tiebreaker. In the 20-dyad sample analyzed here, 5.4% of spurts were marked with uncertainty by a first annotator, and only 1.8% by a second, with a fairly even distribution across strength and polarity levels. This method yielded very high inter-rater agreement between the two annotators. Weighted Cohen's kappas with equidistant penalties were 0.87 for stance strength labels and 0.93 for polarity labels ($p < .001$), with the unweighted kappa for combined labels at 0.88 ($p < .001$).

With the given annotation protocols, uneven distributions across levels were expected, with strong stances particularly rare. **Table 4** shows the distribution of analyzed

Table 3: Stance polarity levels.

Label	Description and examples (applicable only to strength labels 1, 2, 3)
+	Positive: agreement, approval, willing acceptance, encouragement, positive evaluation (e.g., "Sure. Good idea." "Yes! Perfect.")
—	Negative: disagreement, disapproval, rejection, grudging acceptance, hedging, negative evaluation (e.g., "No, I don't think so." "Well, I guess. If you want to.")
(none)	Neutral: none of the above, non-evaluative offering or solicitation of opinions or solutions (e.g., "What should we cut next?" "Let's do this one.")
X	Unclear: cannot be determined.

Table 4: Number of stressed-content vowels by stance strength and polarity.

Strength		Polarity					Total
		Positive (+)	Neutral	Negative (—)	Unclear (x)	NA	
Minimal	(0)					5,630	5,630 (18%)
Weak	(1)	4,622	8,653	217	14		13,506 (42%)
Moderate	(2)	2,040	8,666	1,745	38		12,489 (39%)
Strong	(3)	36	213	167	4		420 (1%)
Total		6,698 (21%)	17,532 (55%)	2,129 (7%)	56 (0%)	5,630 (18%)	32,045

stressed-content vowels (SCVs) by stance strength and polarity, as inherited from the spurts that contained them. Weak and moderate-strength SCVs were similar in proportion, but over half of SCVs were labeled with neutral polarity, a fifth with positive, and very few with negative. Note that vowels with unclear polarity were included in stance strength analysis but removed for polarity analysis; vowels with unclear stance strength were excluded from both analyses.

2.1.3.3. Stance type

Stance type was annotated at a more fine-grained level than stance strength and polarity: Words and phrases were only marked when they performed 'stance acts,' or dialog acts involving stance-taking (Carletta et al., 1997; Fairclough, 2003). Stance act boundaries were determined by the annotators, and acts might divide or span multiple spurts. Both lexical and auditory information was considered when marking a stance act, based on whether the utterance performed the functions shown in **Table 5** within the discourse context. As with stance strength and polarity annotation, annotators listened to both audio channels of a task while annotating one speaker's transcript, and then listened again to annotate the other's. The stance-act type annotation scheme drew on a range of content- and discourse-analytic literature with a variety of stance-related concepts and classifications (Jaffe, 2009), as described below. This resulted in a combination of dimensions that are often examined separately, including elements of persuasion, discourse management, and interpersonal relations, which were combined into one scheme here in order to capture the range of behaviors typical to the collaborative tasks at hand.

Some of the most overt types of stance-taking were included under the *opinion-offering* label (o): evaluation and evaluative description, appraisal, judgment, appreciation, affect/affective stance, assessment, subjectivity, intersubjectivity, positioning, alignment, attitude/attitudinal stance, recommendation, persuasion, modality, modulation, and prediction (Conrad & Biber, 2000; du Bois, 2007; Fairclough, 2003; Hunston & Thompson, 2000; Ogden, 2006).

Table 5: Stance act types.

Label	Description and examples
o	Offer opinion, suggestion (e.g., "I think we should…" "That's really important.")
s	Solicit opinion or agreement (e.g., "What do you think?" "Is that alright?")
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, approval (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; hesitation to offer a stance (e.g., "But that's just me." "Well, I don't know, but…")
t	Teamwork/rapport-building: jokes, teasing, commiseration, comments on tasks
e	Encouragement/praise (e.g., "Good idea." "Now we're getting somewhere!")
i	Strongly-expressive intonation (e.g., incredulous, skeptical, mocking)
x	Unclear (hard to label but clearly stance-related)
b	Backchannels (e.g., "Mm-hm, yeah.")
0	Minimal-stance (stance not clearly present, e.g., factual questions and answers)

In the *convincing/credibility* type (c), speakers engaged in epistemic stance-taking, offering support for their stances by citing knowledge or experience, experts, friends/family, published sources, accepted 'facts,' etc., by explaining their reasoning, or by expressing degrees of commitment, confidence, or certainty (Biber & Finegan, 1989; Conrad & Biber, 2000; Fairclough, 2003; Hunston & Thompson, 2000). *Hedging, softening, or hesitation to offer a stance* (f) could be considered a type of epistemic stance which expresses the converse of the credibility moves in (c), i.e., by showing a lack of commitment, confidence, or certainty in one's own stance. It could also be used for interpersonal stance, e.g., to show deference to another's preferences or authority (Hunston & Thompson, 2000).

In *soliciting another's stance* (s), speakers engaged in both knowledge exchange (Fairclough, 2003) and interpersonal stance-taking, which involved negotiating their positions and power relationships, showing deference and politeness, and/or controlling the flow of conversation and the weights or attention given to each person's stances (du Bois, 2007; Hunston & Thompson, 2000). Both *teamwork/rapport-building* and *encouragement/praise* (t, e) were interpersonal in nature (du Bois, 2007), with speakers working to bolster their cohesiveness as a team by expressing positive sentiments about their jointly-constructed stances, each other, and themselves as team members.

Agreement and *disagreement* (a, d) can be called second order stances (Kockelman, 2004) in that they take stances in relation to previous stances of any type (Conrad & Biber, 2000; du Bois, 2007; Fairclough, 2003; Ogden, 2006). As a polite form of disagreement, *reluctance to accept a stance* (r) adds a layer of positive interpersonal stance to the rejection of a proposition (du Bois, 2007; Fairclough, 2003; Hunston & Thompson, 2000; Ogden, 2006).

The remaining labels allowed for types of stance that were difficult to name (*strongly expressive intonation, unclear* [i, x]) and those which normally carry little or no stance (*backchannels, minimal-stance* [b, 0]). Although *backchannels* were considered to have no/minimal stance (**Table 2**), they were labeled separately for stance type due to their recognizable discourse function and previously-studied acoustic properties (e.g., Beňuš, Gravano & Hirschberg, 2007), which may serve as a useful basis of comparison against stance-carrying types.

Some of the labels served similar functions which were often more difficult to differentiate during annotation. A distinguishing feature between *agreement* and *opinion-offering* (a, o) was whether the utterance took a new stance (o) or merely showed acceptance/approval of an existing one (a). Similarly, lexically positive *backchannels* (b) like 'yeah, right, okay' could be difficult to distinguish from *agreement/acceptance* (a); here the rule of thumb was whether the speaker took (or attempted to take) the floor (a). (The new turn may continue after the agreement, or if the agreement was the entire turn, the other speaker often began a new turn in response, whereas backchannels generally occurred during another speaker's turn.) While *reluctance to accept* and *hedging* (r, f) could sound similar, *reluctance* usually occurred in response to another's stance to soften or avoid rejection, while *hedging* attempted to soften the force of one's own offer, allowing more room for the other to reject it. *Rapport-building* and *encouragement* (t, e) are very similar concepts, as encouragement could be considered a subtype of rapport. However, they were separated here to allow for potentially strong prosodic differences between the more extreme examples, such as individual esteem-boosting verbal 'pats on the back' (e) versus sarcasm or commiseration (t), which on the surface may appear negative but which served to build solidarity (i.e., "At least we're in the same boat"). Finally, labels for general and intonationally-carried 'stanciness' (x, i) were left underspecified to allow for additional classifications that might emerge in future analyses.

Multiple labels were 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) would be labeled (os). Because stance type annotation is more subjective than stance strength and polarity procedures, all annotations were reviewed and corrected by a second annotator. Any areas of uncertainty or disagreement between the first two annotators were settled by a third. In the 20-dyad sample used here, 5% of acts were marked with uncertainty by a first annotator, and only 1% by a second. Labels receiving greater than 5% initial uncertainty included: *reluctance to accept, disagreement, opinion with reasons, softened opinion, strongly-expressive intonation,* and *unclear* (r, d, co, fo, i, x). Finally, stance acts with automatic transcript alignments which deviated substantially from the audio were marked during annotation. These poor alignments made up a small portion of the recordings (4.3% of acts in the 20-dyad sample), and so they were removed from the current acoustic analysis.

Because stance acts were delimited independent of spurt boundaries, they differed in structure from spurts. On average, stance acts in the sample were shorter than spurts, with a mean length of 3.9 words over 1.3 seconds, compared to 6.4 words in 2.2 seconds for spurts. (The speaking rate was unaffected, at about 3 words per second.) As with spurts, stance acts were longer on average in the Budget task (mean 4.4 words, compared to 3.5 in the Inventory task). These patterns held for both sexes.

The 24 stance type labels and label combinations with at least 100 stressed-content vowel tokens were included in the analyses of stance type presented here (**Table 6**). This helped ensure there were enough tokens with each label for reliable comparisons between types. With over 32,000 total vowels, all types in the annotation scheme (**Table 5**) were represented except *encouragement* (e). **Table 6** shows the total number of stance acts with each label, the mean and standard deviation of the number of words and the number of stressed-content vowels (SCVs) per act type, and the total number of SCVs with each label. The most frequent stance act types were *opinion-offering, convincing/reasoning,* and *agreement* (labels o, c, a); together, these comprised 54% of the measured stressed-content vowels. Also frequent were vowels in stretches of speech labeled here as *minimal-stance* (labeled 0, 24% of SCVs); these were not considered parts of stance acts, but they were included in acoustic analyses for comparison. *Opinions with solicitation* or *supporting reasons* (os, co) together contributed just under 9% of all SCVs, and the remaining stance types contributed less than 2% each. Stance act types varied substantially in length, with acts involving *convincing* (c, co, cd, ct, cs, cr) being some of the longest, at about 9 words with nearly 4 SCVs on average, those involving *opinion-offers* (o, os, co, ot, fo, do, ao) next with about 6.5 words and 3 SCVs, other types ranging from 2 to 5 words with about 2 SCVs, and *backchannels* tending to be one-word acts.

2.2. Measurements

After transcription, alignment, and annotation were complete, a Praat script automatically measured the f0 and intensity (Hz, dB) of all vowels at every decile of their duration using Praat's autocorrelation and mean energy functions with a window length of 25 ms, f0 range of 50–300 Hz,[1] and dynamic range of 30 dB. Forced-alignments and automatic measurements were not manually corrected, as the very large size of the data

---

[1] The Praat manual (Boersma & Weenink, 2013, Section Intro 4.2) recommends using a pitch range ceiling of 500 Hz for females. As a post hoc check for effects that the 300 Hz ceiling may have had on the present study, a sample of 17 females was remeasured with a pitch ceiling of 500 Hz. Only 2.5% of their vowel midpoints had an f0 above 300 Hz. With males unaffected by this change, only an estimated 1.5% of all midpoint f0 measurements may have been affected, and so the corpus was not remeasured for the present study.

Table 6: Stance type labels with >100 stressed-content vowels (SCVs).

	Label and Description	Acts	Mean words/act	SCVs
o	offering opinion/suggestion	3,000	5.9	7,991
0	minimal-stance (*often not acts*)	3,427	7.9	7,569
c	convincing/reasoning	1,564	8.7	5,720
a	agreement	3,292	1.9	3,663
os	offer+solicit ("How about…?")	703	5.3	1,786
co	opinion with reasons	267	9.0	1,064
s	soliciting opinion	393	3.7	506
ot	opinion with rapport	137	7.0	386
f	softening/hesitation	345	3.0	378
cd	disagreement with reasons	92	9.6	369
t	teamwork/rapport-building	158	4.9	363
ct	reasons supporting rapport ("That's why we're so good!")	90	8.1	319
ac	agreement with reasons	82	8.1	296
cs	soliciting with reasons ("You think so because…?")	78	7.9	253
x	unclear but stance-related	188	2.7	228
fo	softened offer	89	5.8	216
r	reluctance to accept a stance	184	2.2	173
b	backchannels	193	1.1	146
do	disagreement with alternative	45	8.0	139
i	strongly-expressive intonation	80	2.2	120
at	agreement with rapport	72	3.2	115
d	disagreement	95	3.8	113
cr	reluctance with reasons	28	9.5	111
ao	agree and offer a new opinion	38	5.9	109

set minimized the effects of alignment and measurement errors. However, spurts with very poor alignments were marked during annotation and excluded from analysis; in the current sample, this resulted in excluding about 3.5% of vowels.

Measurements were normalized within-speaker to allow for cross-speaker comparisons. Vowel f0 and intensity were each z-score normalized using the means and standard deviations of all a speaker's measurements taken over all words in both tasks combined. Similarly, vowel duration was z-score normalized within speaker but also within vowel quality to account for intrinsic vowel duration differences (Peterson & Lehiste, 1960; Tauberer & Evanini, 2009). Each vowel's stance strength, polarity, and type labels were inherited from the spurt or stance act to which the vowel belonged. For example, if an utterance of "I agree absolutely" were a spurt marked with moderate strength and positive polarity, and also marked as an act of agreement, the acoustic measurements of each vowel in the utterance would be tagged with 2+ and *agreement*.

3. Results

Signals of stance strength, polarity, and type were found in the duration, fundamental frequency, and intensity of lexically-stressed vowels within content words (stressed-content vowels, SCVs). Because initial analyses showed f0 and intensity patterns holding across vowel duration, the statistics reported below are for these measures at vowel midpoint. A principal components analysis of the *z*-score normalized measures revealed that f0 and intensity aligned with one component which accounted for about half the variance in stance labels, and vowel duration aligned with a second component which accounted for another third of the variance (see Freeman, 2015 for the full analysis).

The primary results reported below come from linear mixed-effects models for each dependent measure (midpoint f0, midpoint intensity, vowel duration) with stance strength, polarity, and type as fixed effects, speaker as a random effect (random intercept), and a random slope for stance strength within speaker. Models with a random slope for stance type failed to converge, as did models with random slopes for both stance strength and polarity together. Results from models with a random slope for polarity are noted for each measure below. Models were computed in R (R Core Team, 2017) using the lme4 package's lmer function (Bates et al., 2015) and the afex package's Satterthwaite estimations to compute *p*-values (Singmann et al., 2019). The smoothing-spline ANOVA plots for each measure were created using the ggplot2 package (Wickham, 2009).

3.1. Fundamental frequency

Fundamental frequency (f0) at vowel midpoint was systematically related to stance strength and type. **Table 7** shows the results for a linear mixed-effects regression model (LMER) with a random slope for stance strength. Mean midpoint SCV f0 was significantly affected by stance strength, with stronger stances successively higher in f0 but no significant difference between minimal-stance and low-strength vowels (labels 0, 1). Several stance types differed from minimal-stance (label 0), as indicated by the stars in **Table 7**. Results from an LMER with a random slope for polarity were nearly identical, and likelihood ratio tests showed that both models provided better fits than one without random slopes ($\chi^2(9) = 85.28$ against LMER with random slope for strength; $\chi^2(5) = 101.70$ against LMER with random slope for polarity, both $p < .001$).

With high overlap between stance types, it was difficult to identify clusters of stance types based on f0 at vowel midpoint. However, Welch's *t* tests identified a few types that were distinct from the others: *Reluctance to accept a stance* (r) and *strongly-expressive intonation* (i) were indistinguishable with the highest f0, *backchannels* (b) had the lowest, and *agreement* (a) dipped from moderate to low ($p < .05$). These relationships can be seen in the smoothing-spline ANOVA plot in **Figure 1**, which shows a contour connecting mean f0 for each stance type cluster at each decile of vowel duration (Gu, 2002; Wassink & Koops, 2013). While f0 generally declined over vowel duration, *agreement* and *backchannels* (a, b) showed sharper slopes. These patterns held in words at all utterance locations, with f0 generally declining over utterance duration.

3.2. Intensity

Stance strength and type were also reliably signaled by intensity at vowel midpoint. **Table 8** shows the results for a linear mixed-effects regression model (LMER) with a random slope for stance strength. Similar to f0, mean midpoint intensity was significantly affected by stance strength, with stronger stances successively higher in intensity but little difference between minimal-stance and low-strength vowels (labels 0, 1). This was influenced by the large number of vowels in weak positive utterances (label 1 +), which

Table 7: Summary of effects in a linear mixed-effects regression model predicting f0 with stance strength, polarity, and type as fixed effects, speaker as a random effect, and a random slope for stance strength within speaker.

Fixed effects	Estimate	SE	t	p
(Intercept)	0.08	0.04	1.84	
strength 1	−0.01	0.04	−0.18	
strength 2	0.13	0.04	3.59	***
strength 3	0.50	0.11	4.74	***
polarity +	−0.02	0.04	−0.57	
polarity 0	−0.01	0.03	−0.38	
type a (agreement)	−0.38	0.03	−11.97	***
type ac (agree+reason)	−0.27	0.07	−3.77	***
type ao (agree+opinion)	−0.24	0.12	−2.02	*
type at (agree+rapport)	−0.16	0.11	−1.46	
type b (backchannel)	−0.61	0.11	−5.78	***
type c (reasoning)	−0.29	0.03	−11.12	***
type cd (disagree+reason)	−0.22	0.07	−3.17	**
type co (opinion+reason)	−0.25	0.04	−5.80	***
type cr (reluctance+reason)	−0.21	0.11	−1.87	
type cs (solicit+reason)	−0.06	0.08	−0.82	
type ct (rapport+reason)	−0.05	0.07	−0.69	
type d (disagree)	−0.09	0.12	−0.74	
type co (disagree+alternate)	−0.22	0.10	−2.29	*
type f (hesitation)	−0.21	0.07	−3.00	**
type fo (offer+hesitation)	0.07	0.09	0.79	
type i (strong intonation)	0.51	0.12	4.11	***
type o (opinion offer)	−0.21	0.02	−8.76	***
type os (offer+solicit)	−0.01	0.03	−0.24	
type ot (offer+rapport)	−0.12	0.07	−1.81	
type r (reluctance)	0.34	0.10	3.53	***
type s (solicit opinion)	0.03	0.06	0.54	
type t (rapport)	0.06	0.08	0.79	
type x (unclear)	0.04	0.13	0.29	
Random effects	**Variance**	**SD**		
Group: speaker				
(Intercept)	0.02	0.15		
strength 1	0.03	0.16		
strength 2	0.02	0.15		
strength 3	0.14	0.38		

Intercept values: strength 0, polarity −, type 0. *** $p < .001$, ** $p < .01$, * $p < .05$.

had lower intensity than minimal-stance and other weak-stance vowels (labels 0, 1, 1–). Polarity levels did not differ substantially in intensity. Several stance types differed from minimal-stance (label 0), as indicated by the stars in the table. Estimates from an
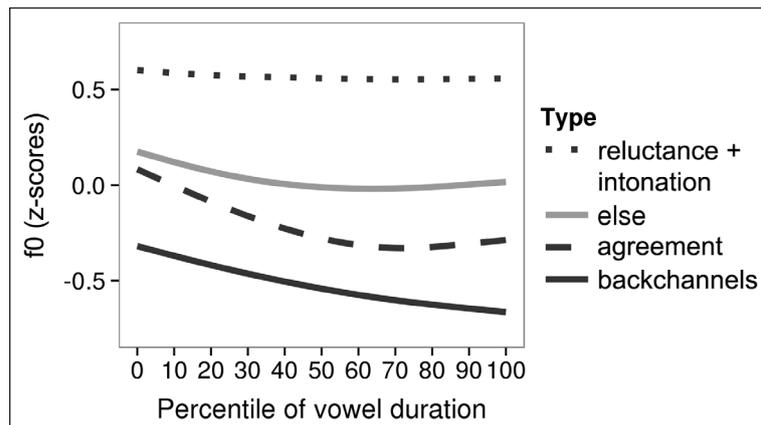
Figure 1: F0 contours by stance type. Contours connect mean values at each decile of vowel duration. "Reluctance + intonation" = combined types: *reluctance to accept a stance* and *strongly-expressive intonation*, which were indistinguishable, as were all other types not listed separately, combined here as "else."

LMER with a random slope for polarity were very similar, but five fixed effects differed in significance between the two (indicated with exclamation points in **Table 8**): Neutral polarity (label 0) and unclear stance (type x) did not have significant effects, but low-strength (label 1), backchannels, and disagreement with reasons (types b, cd) reached significance ($p < .05$). Likelihood ratio tests showed that both models provided better fits than one without random slopes ($\chi^2(9) = 369.41$ against LMER with random slope for strength; $\chi^2(5) = 88.04$ against LMER with random slope for polarity, both $p < .001$).

As with f0, there was high overlap between stance types, but Welch's *t* tests identified a few distinct types. *Agreement with rapport* (at) had the highest intensity and differed significantly from all other types except *strongly-expressive intonation* (i) ($p < .01$), and its intensity dropped less at the ends of utterances than in other types. *Stance-softening or hesitation* (f) had the lowest intensity and overlapped only with *backchannels* (b), the next highest, which in turn overlapped with the next highest, *agreement* (a) ($p < .05$). Both *agreement* and *backchannels* (a, b) dropped more sharply over vowel duration than other types. All other types overlapped heavily and were not clearly distinguishable based on intensity at vowel midpoint. These patterns can be seen in the smoothing-spline ANOVA plot in **Figure 2**, which shows a contour connecting mean intensity at each decile of vowel duration for each stance type cluster. While intensity generally declined over vowel duration (with drops at the edges, as expected near flanking consonants or silence), *agreement* and *backchannels* (a, b) showed sharper slopes, similar to their pattern for f0. The patterns held in words at all utterance locations, with intensity generally declining over utterance duration.

3.3. Vowel duration

Finally, distinctions in stance were also associated with systematic differences in vowel duration. **Table 9** shows the results for a linear mixed-effects regression model (LMER) with a random slope for stance strength, which provided a better fit than a model without a random slope ($\chi^2(9) = 68.17$, $p < .001$ by likelihood ratio test). (An LMER with a random slope for polarity failed to converge.) Strength levels did not differ reliably, but most stance types differed from minimal-stance (label 0), as indicated by the stars in **Table 9**. The results for polarity were less clear. The LMER indicated minimal differences between polarity labels but a distinction between neutral and negative. However, mean SCV duration for positive utterances was longest, 121 ms, compared to 96 ms for negative and 94 ms for neutral stances, and post-hoc Welch's *t* tests showed that positive stances

Table 8: Summary of effects in a linear mixed-effects regression model predicting intensity with stance strength, polarity, and type as fixed effects, speaker as a random effect, and a random slope for stance strength within speaker.

Fixed effects	Estimate	SE	t	p
(Intercept)	0.55	0.04	15.64	***
polarity +	0.03	0.02	1.77	
polarity 0 (!)	0.04	0.02	2.09	*
strength 1 (!)	−0.04	0.03	−1.38	
strength 2	0.13	0.04	3.56	***
strength 3	0.46	0.08	5.55	***
type a (agreement)	−0.07	0.02	−4.30	***
type ac (agree+reason)	0.06	0.04	1.70	
type ao (agree+opinion)	−0.03	0.06	−0.42	
type at (agree+rapport)	0.41	0.06	6.56	***
type b (backchannel) (!)	−0.08	0.05	−1.50	
type c (reasoning)	0.01	0.01	0.75	
type cd (disagree+reason) (!)	0.07	0.04	1.91	
type co (opinion+reason)	−0.03	0.02	−1.37	
type cr (reluctance+reason)	−0.01	0.06	−0.21	
type cs (solicit+reason)	0.01	0.04	0.26	
type ct (rapport+reason)	0.24	0.04	6.42	***
type d (disagree)	0.00	0.06	−0.03	
type co (disagree+alternate)	0.06	0.05	1.15	
type f (hesitation)	−0.18	0.03	−5.35	***
type fo (offer+hesitation)	0.01	0.04	0.18	
type i (strong intonation)	0.28	0.07	4.18	***
type o (opinion offer)	−0.03	0.01	−2.31	*
type os (offer+solicit)	0.09	0.02	4.94	***
type ot (offer+rapport)	0.03	0.03	0.99	
type r (reluctance)	0.08	0.05	1.55	
type s (solicit opinion)	0.06	0.03	1.98	*
type t (rapport)	0.12	0.04	2.95	**
type x (unclear) (!)	−0.13	0.06	−2.14	*

Random effects Group: speaker	Variance	SD		
(Intercept)	0.03	0.19		
strength 1	0.03	0.19		
strength 2	0.04	0.21		
strength 3	0.14	0.38		

Intercept values: strength 0, polarity −, type 0. (!):differs in significance from LMER with random slope for stance polarity. *** $p < .001$, ** $p < .01$, * $p < .05$.

had longer stressed vowel durations than negative and neutral (both $p < .001$), which did not differ. Thus, there may be strong individual differences in the use of vowel duration to signal polarity.
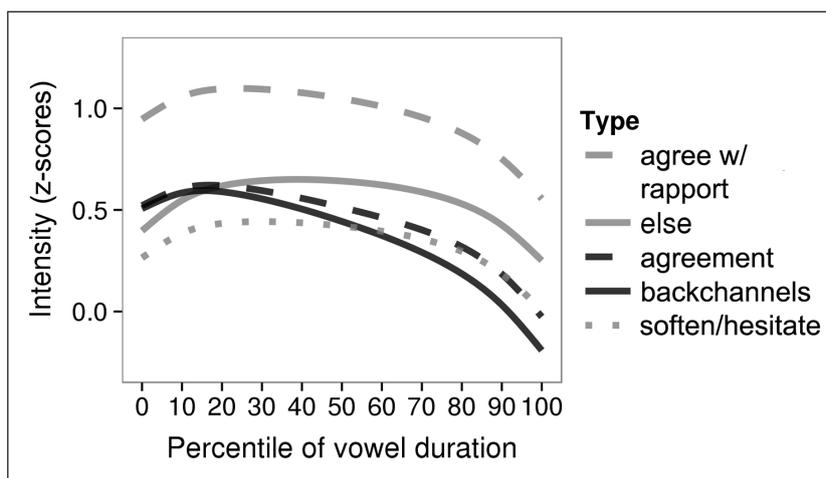
Figure 2: Intensity contours by stance type. Contours connect mean values at each decile of vowel duration.

For stance type, there was again high overlap between types, but Welch's *t* tests identified a few types that differed from most others: *backchannels, agreement with rapport,* and *strongly-expressive intonation* (b, at, i) had some of the longest vowel durations and were only indistinguishable from each other and *unclear stance* (x), which also overlapped *agreement* (a) and five other types. *Agreement* (a) also had longer vowel durations and was only indistinguishable from *unclear* (x) and two other types (fo, r). Other types overlapped heavily and were not clearly distinguishable based on vowel duration.

3.4. *Combined prosodic patterns*

Following the patterns of each measure above, a few of the stance types were differentiated with a combination of prosodic features. *Agreement* (a), one of the most frequent types, showed longer vowel duration and moderately low f0 and intensity which both dipped over the durations of stressed-content vowels. *Backchannels* (b), one of the least frequent types in the corpus, also showed long vowel duration and low-dropping intensity, but their f0 remained low throughout vowel duration. *Reluctance to accept a stance* (r) and *strongly-expressive intonation* (i), also infrequent, showed high f0, the latter also with long vowel duration. *Agreement with rapport* (at) stood out with the highest intensity and longest vowel duration, and *stance-softening/hesitation* (f) showed the lowest intensity.

The same prosodic measures also combined to help differentiate levels of stance strength and polarity. Successively increasing levels of strength were best distinguished by increases in both f0 and intensity, while positive polarity was signaled by longer vowel duration. In combining all three measures, weak-positive utterances (1 +) stood out as having the longest vowels with the lowest f0 and intensity; this group showed the same patterns as the *agreement* type mentioned above (a), as the majority (66%) of *agreeing* stance acts (a) occurred in weak-positive utterances (1 +), and nearly half (47%) of vowels in weak-positive utterances (1 +) contributed to *agreement* (a), with another 5% involved in a combination of types which included *agreement* (ac, ae, aet, af, afo, ai, ao, ar, as, at).

4. Discussion

4.1. *Summary of results*

In this study of a large sample of over 32,000 stressed vowels in content words said by 40 speakers, prosodic measures were found to signal stance strength, polarity, and type. F0 and intensity were most associated with differences in stance strength and type: Both increased with stance strength, and they helped distinguish several stance-act types. *Reluctance to accept* and *strongly-expressive intonation* (r, i) had very high f0, *backchannels*

Table 9: Summary of effects in a linear mixed-effects regression model predicting vowel duration with stance strength, polarity, and type as fixed effects, speaker as a random effect, and a random slope for stance strength within speaker.

Fixed effects	Estimate	SE	t	p
(Intercept)	0.17	0.03	4.95	***
polarity +	−0.01	0.02	−0.38	
polarity 0	−0.05	0.02	−2.13	*
strength 1	0.03	0.03	1.15	
strength 2	−0.01	0.02	−0.65	
strength 3	0.03	0.07	0.38	
type a (agreement)	0.41	0.02	20.10	***
type ac (agree+reason)	−0.19	0.05	−4.02	***
type ao (agree+opinion)	−0.08	0.08	−1.06	
type at (agree+rapport)	0.60	0.08	7.50	***
type b (backchannel)	0.85	0.07	12.53	***
type c (reasoning)	−0.21	0.02	−12.15	***
type cd (disagree+reason)	−0.25	0.05	−5.38	***
type co (opinion+reason)	−0.23	0.03	−8.15	***
type cr (reluctance+reason)	−0.17	0.08	−2.15	*
type cs (solicit+reason)	−0.14	0.05	−2.65	**
type ct (rapport+reason)	−0.10	0.05	−2.09	*
type d (disagree)	−0.22	0.08	−2.81	**
type co (disagree+alternate)	−0.21	0.07	−3.09	**
type f (hesitation)	−0.08	0.04	−1.86	
type fo (offer+hesitation)	−0.02	0.06	−0.31	
type i (strong intonation)	0.23	0.09	2.71	**
type o (opinion offer)	−0.17	0.02	−10.34	***
type os (offer+solicit)	−0.16	0.02	−6.64	***
type ot (offer+rapport)	−0.20	0.04	−4.45	***
type r (reluctance)	0.14	0.06	2.23	*
type s (solicit opinion)	−0.11	0.04	−3.00	**
type t (rapport)	−0.09	0.05	−1.72	
type x (unclear)	0.25	0.08	3.27	**

Random effects Group: speaker	Variance	SD
(Intercept)	0.02	0.14
strength 1	0.01	0.12
strength 2	0.00	0.05
strength 3	0.05	0.23

Intercept values: strength 0, polarity −, type 0. *** $p < .001$, ** $p < .01$, * $p < .05$.

(b) very low, and *agreement* (a) low-dipping; the latter two also showed sharply-dropping intensity, with *backchannels* lower overall. *Stance-softening/hesitation* (f) showed the lowest intensity and *rapport-building agreement* (at) the highest. While most of these types

also had longer vowels, vowel duration did not reliably differentiate them. While positive polarity showed longer vowel duration, individual differences between speakers may cloud the use of duration in signaling polarity. Finally, *weak-positive agreement* (a,1 +) stood out with the longest vowels and lowest f0 and intensity. **Table 10** summarizes these results.

These findings support the prediction that information about stance is carried in prosodic features of the acoustic speech signal. It stands to reason that variations in prosody play a strong role in conveying the many complex and subtle meanings of opinions and attitudes. At a phrasal level, many well-known intonational contours can be overlaid on identical lexical/syntactic material to change the meaning from statement to question, scolding to incredulous, genuine to sarcastic, and so on, but in naturally-occurring speech, such well-defined tunes are affected by a host of other contextual factors, making it more difficult to tease apart the acoustic components that contribute to each aspect. This study identified some components of stance meanings as they were carried on stressed vowels in content words, and while phrasal-level analysis is certainly called for in future work, the very large sample size used here allows pieces of the broader pattern to emerge. Again, it stands to reason that stronger stances had higher f0 and intensity, with increased effort during delivery indicating greater investment; that backchannels and weak agreement were quiet and low-pitched; that rapport-building agreement was delivered energetically; that downplaying a stance was done quietly; and that complex stances (e.g., reluctance to accept an idea without outright rejection) carried complex intonation patterns. Such findings form a solid foundation for expansion into both broader and more detailed acoustic investigations.

4.2. Limitations

As some of the first work to report acoustic signals of stance-taking, this study had several limitations, including a 'flattening' of the prosodic information in a spurt caused by collapsing vowel measurements across all spurt positions. Local speaking rate, lexical frequency, and predictability in context were also not considered in detail. Finally, other types of spoken interaction are likely to involve stance types or prosodic contours that are not well represented in the collaborative tasks used here, which encouraged cooperation with low stakes and no consequences attached to any decision the participants made. More competitive tasks or controversial topics are likely to elicit more disagreement, persuasion, and stronger opinions, which may be expressed with distinct prosodic cues.

Table 10: Summary of results.

	Stance feature/type	F0	Intensity	Duration
	Strength	increases with strength levels	increases with strength levels	—
	Polarity	—	—	unclear
r; i	reluctance to accept a stance; strongly-expressive intonation	very high	—	long
at	agreement with rapport	—	very high	very long
a,1+	weak-positive agreement	low-dipping	dropping	long
b	backchannels	very low	low-dropping	long
f	softening/hesitation	—	low	—

4.3. Conclusion

This study provides an initial sketch of the prosodic cues to stance, the ways in which components like f0, intensity, and duration can be manipulated and combined to send complex messages about our attitudes, opinions, and interpersonal relationships. Such information not only deepens our understanding of human communication but also contributes to the growing body of computational work on sentiment analysis (see e.g., Mäntylä, Graziotin, & Kuutila, 2018), for use in both automatic detection and human-interactive production. Given that many other types of information—social/indexical, discursive, structural, etc.—are sent in the same acoustic stream, stance should be considered as a potential influencing factor when designing and analyzing studies of variation in pronunciation and prosody in natural speech.

Acknowledgements

Competing Interests

The author has no competing interests to declare.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Beňuš, S., Gravano, A., & Hirschberg, J. (2007). The prosody of backchannels in American English. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)* (pp. 1065–1068).

Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text—Interdisciplinary Journal for the Study of Discourse, 9*(1), 93–124. DOI: https://doi.org/10.1515/text.1.1989.9.1.93

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* Harlow, Essex, UK: Pearson Education Ltd.

Biber, D., & Staples, S. (2014). Exploring the prosody of stance. In T. Raso & H. Mello (Eds.), *Spoken corpora and linguistic studies* (pp. 271–294). Philadelphia, PA: John Benjamins. DOI: https://doi.org/10.1075/scl.61.10bib

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [computer program], version 5.3.55. Retrieved from http://www.fon.hum.uva.nl/praat/

Carletta, J., Isard, S., Doherty-Sneddon, G., Isard, A., Kowtko, J. C., & Anderson, A. H. (1997). The reliability of a dialogue structure coding scheme. *Computational Linguistics, 23*(1), 13–31.

Conrad, S., & Biber, D. (2000). Adverbial marking of stance in speech and writing. In S. Conrad & D. Biber (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 56–73). New York, NY: Oxford University Press.

Dehé, N., & Wichmann, A. (2010). Sentence-initial I think (that) and I believe (that): Prosodic evidence for use as main clause, comment clause and discourse marker. *Studies in Language, 34*(1), 36–74. DOI: https://doi.org/10.1075/sl.34.1.02deh

Du Bois, J. W. (2007). The stance triangle. In R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 139–184). Amsterdam, Netherlands: John Benjamins. DOI: https://doi.org/10.1075/pbns.164.07du

Englebretson, R. (2007). Stancetaking in discourse: An introduction. In R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction* (pp. 1–26). Amsterdam, Netherlands: John Benjamins. DOI: https://doi.org/10.1075/pbns.164

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research.* New York, NY: Routledge. DOI: https://doi.org/10.4324/9780203697078

Freeman, V. (2014). Hyperarticulation as a signal of stance. *Journal of Phonetics, 45,* 1–11. DOI: https://doi.org/10.1016/j.wocn.2014.03.002

Freeman, V. (2015). The phonetics of stance-taking. Doctoral dissertation. Seattle, WA: University of Washington Press.

Freeman, V., Chan, J., Levow, G.-A., Wright, R., Ostendorf, M., & Zayats, V. (2014). Manipulating stance and involvement using collaborative tasks: An exploratory comparison. *Proceedings of INTERSPEECH 2014, the 15th annual conference of the international speech communication association* (pp. 2238–2242). ISCA Archive: http://www.isca-speech.org/archive/interspeech_2014

Freeman, V., Levow, G.-A., Wright, R., & Ostendorf, M. (2015). Investigating the role of 'yeah' in stance-dense conversation. *Proceedings of INTERSPEECH 2015, the 16th Annual Conference of the International Speech Communication Association* (pp. 3076–3080). ISCA Archive: http://www.isca-speech.org/archive/interspeech_2015

Freese, J., & Maynard, D. W. (1998). Prosodic features of bad news and good news in conversation. *Language in Society, 27*(2), 195–219. DOI: https://doi.org/10.1017/S0047404500019850

Gu, C. (2002). *Smoothing spline ANOVA models.* New York, NY: Springer. DOI: https://doi.org/10.1007/978-1-4757-3683-0

Haddington, P. (2004). Stance taking in news interviews. *SKY Journal of Linguistics, 17,* 101–142. Available online: http://www.linguistics.fi/skyjol-en.shtml

Hunston, S., & Thompson, G. (2000). Evaluation: An introduction. In S. Hunston & G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 1–27). New York, NY: Oxford University Press.

Jaffe, A. (2009). Introduction: The sociolinguistics of stance. In A. Jaffe (Ed.), *Stance: Sociolinguistic perspectives* (pp. 3–28). New York, NY: Oxford University Press.

Kockelman, P. (2004). Stance and subjectivity. *Journal of Linguistic Anthropology, 14*(2), 127–150. DOI: https://doi.org/10.1525/jlin.2004.14.2.127

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology and sound change.* Berlin, Germany: Walter de Gruyter. DOI: https://doi.org/10.1515/9783110167467

Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review, 27,* 16–32. DOI: https://doi.org/10.1016/j.cosrev.2017.10.002

Morgan, N., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Janin, A., Pfau, T., Shriberg, E., Stolcke, A. (2001). The meeting project at ICSI. *Proceedings of the first international conference on human language technology research.* DOI: https://doi.org/10.3115/1072133.1072203

Ogden, R. (2006). Phonetics and social action in agreements and disagreements. *Journal of Pragmatics, 38*(10), 1752–1775. DOI: https://doi.org/10.1016/j.pragma.2005.04.011

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* (JASA), *32*(6), 693–703. DOI: https://doi.org/10.1121/1.1908183

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2019). afex: Analysis of factorial experiments. R package version 0.23-0. https://CRAN.R-project.org/package = afex

Tauberer, J., & Evanini, K. (2009). Intrinsic vowel duration and the post-vocalic voicing effect. *Proceedings of INTERSPEECH 2009, the 10th annual conference of the international speech communication association.* ISCA Archive: https://www.isca-speech.org/archive/archive_papers/interspeech_2009

U.S. Census Bureau. (2010). State and county quickfacts: King County, Washington. Retrieved from http://www.census.gov/quickfacts/table/POP060210/53033

Ward, N. G., Carlson, J. C., & Fuentes, O. (2018). Inferring stance in news broadcasts from prosodic-feature configurations. *Computer Speech & Language, 50,* 85–104. DOI: https://doi.org/10.1016/j.csl.2017.12.007

Wassink, A. B., & Koops, C. (2013). Quantifying and interpreting vowel formant trajectory information. *Paper presented at New Ways of Analyzing Variation* (NWAV 42).

Wickham, H. (2009). *Elegant graphics for data analysis.* Springer-Verlag New York.

Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics '08*. DOI: https://doi.org/10.1121/1.2935783