The Phonetics of Stance-taking

Valerie Freeman

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Richard Wright, Chair

Gina-Anne Levow

Betsy Evans

Program Authorized to Offer Degree:
UW Department of Linguistics

University of Washington

Abstract

The Phonetics of Stance-taking

Valerie Freeman

Chair of the Supervisory Committee:
Professor Richard Wright
Department of Linguistics

Stance – attitudes and opinions about the topic of discussion – has been investigated textually in conversation- and discourse analysis and in computational models, but little work has focused on its acoustic-phonetic properties. It is a challenging problem, given the complexity of stance and the many other types of meaning that must share the same acoustic channels, all of which are overlaid on the lexical and syntactic material of the message. With the goal of identifying automatically-extractable, acoustically-measurable correlates of stance-taking, this dissertation presents a new audio corpus of stance-dense interaction and three phonetic experiments which find signals of stance in prosodic measures of pitch, intensity, and duration. The ATAROS corpus contains pairs of speakers engaged in collaborative conversational tasks designed to elicit frequent changes in stance at varying levels of involvement. Interactions are transcribed, time-aligned to the audio, and manually annotated for stance strength (none, weak, moderate, strong), polarity (positive, negative, neutral), and stance type (e.g., opinion-offering and soliciting, (dis)agreement, persuasion, rapport-building, etc.). In the first experiment, combinations of pitch and intensity contours are shown to differentiate four discourse functions within a small sample of instances of the word ‘yeah’ that contribute to negative stances. In the second experiment, vowel duration and intensity separate six common stance-act types in over 2200 ‘yeahs,’ changes in pitch and intensity correlate with stance strength, and all three measures are involved in signaling positive stance. The third and largest experiment examines over 32,000 stressed vowels in content words spoken by 40 speakers and finds that pitch and intensity increase with stance strength, longer vowel duration is the primary signal of positive polarity, and a combination of these measures helps distinguish several notable stance-act types, including: agreement in general, weak-positive agreement, rapport-building agreement, reluctance to accept a stance, stance-softening, and backchannels. These results, and the corpus itself, contribute to the study and understanding of the acoustic-phonetic properties of the social and attitudinal messages conveyed in natural speech, information which may be of use to future work in theoretical, experimental, and computational linguistics.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

Table Number Page

GLOSSARY

ATAROS: [əˈtaɹoʊs] Abbreviation for project titled "Automatic Tagging and Recognition of Stance," funded by NSF IIS #1351034, which created the audio ATAROS corpus of stance-dense collaborative conversation (Chapter 2) on which the studies reported here are based. PIs: Drs. Gina-Anne Levow, Richard Wright, Mari Ostendorf, Departments of Linguistics and Electrical Engineering, University of Washington. Related publications (including [28, 29, 30, 31, 33, 61, 62]) and access to the corpus for research purposes can be found through the Linguistic Phonetic Lab website: `http://depts.washington.edu/phonlab/projects.htm`

DIALOG ACT: A topic-specific speech act; often used in annotation in spoken dialog systems (e.g., [14, 92]).

DYAD: A pair of speakers.

SPEECH ACT: The performative function of an utterance, e.g., declarative, question, greeting [1, 23, 86].

SMOOTHING-SPLINE ANOVA (SSANOVA) PLOT: Plots of smooth splines connecting mean measurement values across measurement time points, often with confidence intervals around the means [42]. Splines resemble pitch/intensity/formant traces on a spectrogram and can be used to compare contour shapes and areas of overlap [100] (see Experiments 1-3, Chapters 4-6).

SPURT: Stretch of speech said by one speaker between at least 500 ms of silence. Manually marked during transcription (Section 2.3.1) and used as the time-unit of annotation for stance strength and polarity (Section 2.3.2).

STANCE: Attitudes, opinions, beliefs, or judgments about an object, person, or proposition relevant to the topic of discussion [8, 20, 43]. Most common related terms: evaluation [20, 43, 50], sentiment and subjectivity [75, 104]. See Sections 1.1 and 2.3.3.

STANCE-TAKING: Discourse activity of expressing stance [8, 20, 43], Section 1.1.

STANCEY/STANCINESS: (working terms used within the ATAROS team) Quality of carrying stance, applied to units of speech, speakers, interactions; gradations denote quantity (frequency) or strength, e.g., “very stancey, stancier” could indicate a high(er) number/frequency of stance moves and/or strong(er) stance strength.

STANCE ACT: Dialog act involving stance-taking (opinion-offering, (dis)agreement, convincing, etc.). Used as the time-unit of stance type annotation (Section 2.3.3, Appendix E).

STANCE POLARITY: Classification (positive, negative, neutral) applied to spurts during annotation (Section 2.3.2, Appendix D).

STANCE STRENGTH: Within-speaker classification (none, weak, moderate, strong) applied to spurts during annotation (Section 2.3.2, Appendix D).

STANCE TYPE: Category of stance acts (e.g., opinion-offering, soliciting, accepting, softening) identified and labeled during annotation (Section 2.3.3, Appendix E).

STRESSED-CONTENT VOWELS: Lexically stressed (primary or secondary) vowels in content words; the focus of most analyses in Experiment 3 (Chapter 6).

ACKNOWLEDGMENTS

First and foremost, I would like to thank Dr. Richard Wright, most excellent advisor, teaching and research supervisor, for his guidance and unfailing support throughout my time as a graduate student. I hope you know how much the time and energy you invest in your students is appreciated. I couldn't ask for a better mentor, and I only hope I can live up to your example, as an advisor, colleague, and productive researcher.

Special thanks to Drs. Gina-Anne Levow and Mari Ostendorf for providing strong models of cross-disciplinary cooperation and collegiality, even in the frenzy to meet publication deadlines. I have truly felt a valued member of the research team under your leadership.

Thanks especially to Heather Morrison for her invaluable assistance with corpus collection, transcription, and annotation, and to the transcribers and annotators who have contributed significantly to the corpus: Lauren Fox, Phoebe Parsons, Nicole Chartier, Marina Oganyan, Max Carey, and Andrew Livingston.

Thanks to fellow RAs Yi Luan and Julian Chan for creating such useful tools and patiently teaching me how to use them, and to the rest of the computational linguistics team, Trang Tran, Alena Hrynkevich, Sam Tisdale, Victoria Zayats, and Maria Antoniak, for their contributions to corpus analysis.

Thanks also to John Riebold, Dan McCloy, and Rachael Tatman for their technical and statistical expertise, to Drs. Betsy Evans and Alicia Wassink for guidance along the way, to Mike Furr for keeping track of the hodgepodge of finances, and to members of the UW Phonetics and Sociolinguistics Lab groups, seminars, and conference audiences for their feedback and encouragement as the project progressed.

Last but not least, thanks to Mark Ellison for his daily support, outside perspectives,

and knowing when to let me work and when to make me quit for the day.

Portions of this work were supported by National Science Foundation grant IIS 1351034, National Institute of Health grant R01 DC60014, and the University of Washington Department of Linguistics Excellence in Linguistic Research Graduate Award. Portions have been presented at recent conferences and reported in proceedings and technical reports, as indicated in footnotes at the beginning of each chapter.

Chapter 1

INTRODUCTION

Stance, or a speaker's attitudes and opinions, can be conveyed in many ways, but with only a fraction of the message sent through the words themselves, much of the information must be present in the delivery, the speech signal itself. Just as changes in pronunciation and prosody can transform the meaning of a sentence from statement to question, similar changes can affect the transmission and reception of messages about many levels of social and attitudinal information. While phonetic correlates of information structure, discourse structure, and such social aspects as the region, gender, ethnicity, identity, etc. of speakers – and perceptions and interpretations of these features by listeners – have been studied in various sociolinguistic and computational fields, the phonetic properties of stance-taking have received much less attention. This leads to questions of how stance is signaled acoustically. For example, how do we express strong vs. weak opinions, or positive vs. negative attitudes? How do we convey enthusiastic vs. reluctant agreement, take confident vs. uncertain positions, engage in persuasion or show deference, all without changing the words we use? This dissertation addresses such questions by presenting some of the first work to find acoustically-measurable correlates of stance-taking in natural speech. The main contributions of the work[1] include the design and creation of a large audio corpus of stance-dense collaborative conversation and three phonetic experiments in which automatically-extracted acoustic-prosodic measures are found to signal aspects of stance type, strength, and polarity.

[1] Portions of this work were funded by NSF IIS 1351034, NIH R01 DC60014, and the University of Washington Department of Linguistics and have been presented at several conferences, indicated in footnotes at the beginning of each chapter. Information in this chapter also appears in [28, 29].

Building a corpus and annotation schema specifically for stance-related research improves upon difficulties in past work with existing corpora, including the abilities to elicit a high density of stance moves on specified topics, control and manipulate speaker demographics, and begin to disentangle acoustic-prosodic features of stance-expression from other structural, social, and indexical meanings present in the speech signal. By labeling utterances holistically for the strength and polarity of their stances, as well as identifying more specific subtypes of stance acts (opinion-offering, agreement, etc.), two levels of stance features can be examined, both separately and in interaction.

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 [22, 52] for summaries). The work presented here takes a broad view of *stance* as used in discourse- and conversation analytic approaches: "personal feelings, attitudes, value judgments, or assessments" [8, p. 966], and of their expression, the social activity of *stance-taking*, often called *evaluation* [20, 43, 50]. Du Bois [20] describes 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. This is precisely the process elicited in the collaborative tasks designed for this study, detailed in Section 2.2.3.

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 factors. Conversation- and discourse-analytic approaches provide many descriptions of stance, often seated in fine-grained content analysis (e.g., [7, 19, 20, 22, 43, 50, 52]). A precursor to the current project [25, 26] drew on such existing frameworks for stance type classification to identify areas of stance-expression for subsequent phonetic analysis. As some of the first work to focus on acoustic properties of stance-taking, it reports that stance-expressing phrases have faster speaking rates, longer stressed vowels, and more expanded vowel spaces when compared to stance-neutral phrases. However, as a small-scale study, it was only able to examine stance at the coarsest level – binary presence/absence, collapsing many types of stance-taking acts identified in the conversation/discourse-analytic literature. This is addressed in the current studies by separately labeling such stance act types, as well as stratifying utterances more holistically by the strength and polarity of their stances. The stance act types and strength/polarity features are then compared on acoustic-prosodic measures, following the argument that since stance presence is signaled acoustically, its components and subtypes may also display differing acoustic properties. The resulting findings can be of use in both theoretical/experimental linguistics and in automatic recognition research. In this related field, concepts similar to stance are studied under terms such as *sentiment* and *subjectivity*, or types of *private states* [75, 79, 104], with most work focusing on lexical or syntactic cues in annotated corpora (e.g., [75, 89, 104]), finding it difficult to implement or interpret analyses of audio content [48, 70, 81, 90, 105, 106].

The work presented here brings together methods used in multiple areas of study: content analysis of stance-taking as devised in conversation/discourse analysis and often applied in automatic sentiment recognition; measurement of acoustic-prosodic features commonly employed in phonetics; and automatic extraction of a large number of such measurements over an audio corpus of naturalistic interactions, as favored in computational and corpus linguistics. The combined approach leverages advantages of detailed qualitative analysis with quantitative measurement on a scale that provides statistical power, greater potential for generalizability, and applicability to future work on similar and related topics.

1.2 Study structure

In the pursuit of acoustic signals of stance-taking, a new corpus is constructed in order to elicit naturalistic stance-dense conversation with control over speaker demographics, discussion topics, and recording conditions. Chapter 2 describes the design, collection, and structure of this corpus, which includes collaborative tasks designed to elicit different strengths and types of stance-taking. Such features of stance-taking are predicted to display acoustic correlates, as stated in Chapter 3. Next, three experiments are presented using increasingly larger samples of the corpus. Experiment 1 (Chapter 4) focuses on negative uses of the word ‘yeah.’ Although rare, the few instances of ‘negative yeah’ provide a test case for content analysis and acoustic methods subsequently applied to a larger sample of ‘yeahs,’ described in Experiment 2 (Chapter 5). With the examination of stance type, strength, and polarity, this second study serves as a pilot for the main investigation of the acoustic-prosodic features of stance-taking and stance acts throughout the corpus, described in Experiment 3 (Chapter 6). Finally, results, contributions, and future work are discussed in Chapter 7. Elicitation materials and annotation schema used to create the corpus are provided in the appendices.

Chapter 2
THE ATAROS CORPUS

ATAROS [əˈtaɹoʊs], short for “Automatic Tagging and Recognition of Stance,” is a project involving phoneticians, computational linguists, and speech-signal-processing engineers who seek to identify signals of stance-taking in the acoustic speech signal, both for linguistic research and for implementation in automatic stance detection. One of this dissertation’s main contributions to this project is the design and creation of the ATAROS corpus, which is constructed to elicit a high frequency of stance-taking in naturalistic conversation. This chapter[1] describes all aspects of the corpus: motivation and advantages (Section 2.1), design and collection (Section 2.2), transcription and annotation (Section 2.3), structural and acoustic characteristics (Section 2.4).

2.1 Motivation

The ATAROS corpus is designed specifically for the purpose of identifying acoustically-measurable signals of stance-taking in natural speech, and as such, it provides several advantages over speech collected for other purposes. Limitations of existing corpora include issues with recording quality, speaker attributes, and the type or content of speech.

Recording quality varies widely when audio is gathered from sources not created for linguistic analysis. Common concerns are recording conditions and microphone type and placement, which often affect the signal-to-noise ratio and acoustic intensity. For example, corpora such as CALLHOME and SWITCHBOARD [56, 37] were collected over the phone, resulting in band-limited audio. Similarly, intensity is an unreliable measurewhen loudness

[1] Portions of this work were presented at the 2014 Acoustical Society of America Spring Meeting [30] and are reported in the *Proceedings of Interspeech 2014* [28], *ATAROS Technical Report 1* [29], the *Proceedings of SLT 2014* [61], and the *Proceedings of Interspeech 2015* [31].

is adjusted for public broadcast (TV, radio, etc.) or when the distance between a speaker and microphone varies unpredictably. These were problems in a precursor to the ATAROS project [25, 26], in which acoustic cues to stance were analyzed on a televised political talk show.

More specific to the study of linguistic variation is the ability to disentangle within- and between-speaker variation. Factors to consider include speaker demographics, social roles, and the amount and type of speech collected from each person. Social factors such as age, gender, ethnicity, dialect region, and the relationship between speaker and audience commonly correlate with linguistic variation, but these attributes are not always known or controlled in audio collections such as CALLHOME, the ICSI Meeting Corpus and the AMI Corpus [56, 69, 13]. This was also a problem in the political talk show study [25, 26] because the episode under analysis contained only males, each from a different dialect region (which was only possible to determine because they happened to have publicly-available biographic information). This made it difficult to compare regionally-variable features such as vowel space configurations and to determine whether differences in prosodic patterns reflected individual speaker traits or additional (as-yet under-described) regional dialect features.

The type of speech also matters; of interest here is stance in spontaneous, unscripted, naturalistic conversation, which differs from read or performed speech in ways that may affect stance-taking. For example, the personal motives underlying stance moves – and the way they are delivered – may differ greatly between social roles (boss, friend, parent, etc.) and between settings (meetings, public discussion, personal conversation, etc.) [4, 34]. More to the point, many situations do not naturally involve a high density of the phenomenon under investigation. This is particularly relevant for stance-taking, which might be found in high densities in more formal, scripted situations such as debates but less reliably in conversation. Finally, when intra-speaker variation is desired, a larger amount of speech is required from each speaker in each condition predicted to have an effect, in order to obtain enough power for linguistic analysis and to provide sufficient material for computational modeling and machine learning.

All of the above factors are addressed with the creation of the ATAROS corpus. Its high-quality audio recordings are ideal for acoustic analysis, with head-mounted microphones in separate channels and a quiet environment. Conversation is unscripted but focused around collaborative tasks that require increasing levels of involvement and stance-taking. With some structure provided by the tasks, many target words are repeated throughout the recordings, enabling straightforward comparisons within and across both speakers and tasks. Speakers are paired in dyads and complete all tasks in one session, yielding a similar amount of speech in each task from each speaker. Basic demographics are collected and controlled: speakers are matched roughly for age and either matched or crossed by sex, yielding similar numbers of male-male, female-female, and male-female dyads. This arrangement allows for the comparison of speech patterns used by each sex and directed to other speakers of the same or different sex; matching by age helps mitigate differences in speech patterns related to power or politeness dynamics that may be invoked when addressing people of clearly differing ages [4, 11, 34, 91]. All participants are native English-speakers from only one dialect region, the Pacific Northwest. Controlling for dialect region is especially useful in these initial stages of isolating linguistic behavior attributable to stance or involvement without the potential confound of differences between dialects (e.g., vowel inventories, pause durations, pitch patterns, backchannel behavior; [25]).

2.2 Design and collection

Many steps are involved in creating a corpus of the size desired for the current studies, including principled speaker selection, task design and recording procedures.

2.2.1 Speakers

Speakers are recruited from the Seattle area via flyers, online classifieds, listservs, and emails to participants of previous studies conducted at the University of Washington Linguistic Phonetics Lab, where recordings are made. The inclusion criteria detailed below are established during scheduling emails or phone exchanges.

As mentioned in Section 2.1, all speakers are adult native English-speakers from one dialect region, the Pacific Northwest, broadly defined as Washington, Oregon, and Idaho. To qualify for the corpus, speakers must have spent the majority of their childhood in the Northwest, from around age 5 through high school, ideally with less than two years living outside the region during this time, and they must consider English one of their native languages. (Information on other language experience is gathered during the recording session using a brief demographic questionnaire, as described in Section 2.2.3.) This helps control for regional-dialectal differences in pronunciation and prosody; in the future, speakers from other dialect regions may be recorded and analyzed for comparison.

Ethnicity is not controlled, as it has not been a significant factor in prior studies of Northwest English (cf. [84, 85]). However, of the 20 speakers recorded to date who self-identified their ethnic background before their recording sessions, 16 (80%) reported European descent, and 5 (25%) reported Asian-Pacific descent (Japanese, Filipino, Pacific Islander). These proportions are consistent with the general ethnic makeup of the Seattle area [12].

To cover a breadth of the speaker population, speaker ages range from 18 to 75. This upper limit is placed to help reduce the likelihood of significant age-related hearing loss. To qualify, speakers must self-report that they have no history of hearing problems. Speakers with apparent speech impediments or notably odd mannerisms are not disqualified as long as there is no difficulty in understanding their speech; however, they may be excluded from analysis if the disruption is large. (Of those recorded to date, only one speaker has been considered for such exclusion based on a possible impediment or substantial influence from a non-English language; three others have been marked as last priority for processing due to unusual mannerisms, i.e., a frequent use of character voices or sarcasm. None of these speakers appear in the studies reported here.)

Currently, all speaker dyads are made up of strangers matched roughly by age group and either crossed or matched by sex. During scheduling exchanges, age group is requested via ‘decade’ (under 30, 30s, 40s, 50s, 60+), and an effort is made to schedule pairs who are in the same or adjacent decade groups. Also during scheduling, at least one speaker’s name is given to the other, with permission, to ensure the two do not already know each other. Speaker sex may also be requested beforehand in order to balance the makeup of dyads in the corpus. Speakers of a given sex and/or age group may be solicited specifically during recruitment for the same reason (e.g., "We especially need more men and people age 40-60."). These restrictions help control for style factors which are affected by a speaker's audience [4, 34, 91]; future recording phases are expected to include friends/family and dyads of differing ages. See Section 2.4.1 for detailed distributions of speakers and dyads by the above demographic factors.

2.2.2 Recording conditions

Recordings are made in a sound-attenuated booth on the University of Washington campus in Seattle. The booth measures approximately 7 feet by 10 feet and contains a card table, 2-4 chairs, and a small heavy table with a computer screen and keyboard. Each participant is fitted with a head-mounted AKG C520 condenser microphone [36] connected by XLR cable to a separate channel in an M-Audio Profire 610 mixer [64] outside the booth. The mixer is connected to an iMac workstation that uses Sound Studio (version 3.5.7) [88] to create 16-bit stereo WAV-file recordings at a 44.1 kHz sampling rate. The computer screen in the booth is connected to the iMac as a second monitor where instructions are displayed for two of the tasks.

In most cases, the entire recording session (described in Section 2.2.3) is captured in one audio file, but more than one may be created due to long conversations, breaks between tasks, or technical difficulties.

2.2.3 Tasks

After a brief demographic questionnaire, each dyad completes five collaborative problem-solving tasks designed to elicit frequent changes in stance and differing levels of involvement or engagement. There are two groups of tasks, each of which uses a set 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, cf. e.g., [60]). Each group of tasks begins with a find-the-difference list task intended to elicit stance-neutral first-mentions of the items to be used in subsequent tasks. These provide a baseline for comparison against the pronunciations of the same items in subsequent stance-dense tasks, and the isolation of first-mentions of all items helps separate their introduction as new information in the discourse, which has been found to interact with stance-expression [25, 26]. The other three tasks (Inventory, Survival, Budget) are designed to elicit increasing levels of involvement and stronger stances. By providing topics with varying degrees of expected personal interest or investment, speakers are encouraged to express a wider variety of stance types and strengths, including gradations of (dis)agreement, commitment, and negotiation, for example.

The task design builds on a small pilot study [32] in which two male-male and two female-female dyads completed the Inventory and Survival tasks described below. With an adaptation of the stance-annotation scheme used in previous work [25, 26], utterances including the target items discussed in the tasks were tagged for stance-related features such as overt opinion, evaluative description, reasoning, persuasion, and negotiation. Initial analysis considered a greater number of targets with multiple tags to reflect a higher density of stance moves and multiple tags on a target to indicate stronger stance. In the Survival task, 41% of target words were multiply-tagged, 9% with three or more tags, compared to 26% and 1% in the Inventory task. The greater density of stance moves in the Survival task can be related to increased involvement, a connection also supported by an increase in signs of investment, such as extended attempts at persuasion and citing personal experience in support of opinions, behaviors which are also found in the current study. Because both tasks elicited a relatively small proportion of strong stances (triply-tagged targets), the Budget task was added with the intention of further increasing involvement, as financial decisions often evoke strong opinions. However, care was taken to avoid 'hot button' issues which may elicit strong emotional responses, both for the comfort of participants and because strong emotions affect the speech signal in ways that may interact with more moderate stances (cf.

e.g., [103]). Similarly, the stance-neutral tasks were added to balance the scale of intended levels of involvement and to separate the introductions of the items as new information from subsequent mentions as given in the discourse [15, 25, 26, 78].

The following sections describe the ATAROS corpus collaborative tasks in the order they are administered for the current study. Complete elicitation materials can be found in the indicated appendices.

Demographic questionnaire

After speakers are seated in the recording booth, and microphones and sound levels are adjusted, the researcher orally administers a brief demographic questionnaire, noting speakers' responses and later entering them into a secured subject database used in previous and related studies. Each speaker is assigned a unique alpha-numeric ID code that identifies region and sex, e.g. 'NWF025' for the 25th Northwest female in the subject database, so that speakers' names are not used as identifiers in the corpus. The questionnaire asks speakers' age and sex, where they grew up, and what languages they know (see Appendix A for the complete form). For the current study, these questions are used to ensure that all are native speakers of Pacific Northwest English, but the information could be useful to additional sociolinguistic analyses in the future. The procedure also helps speakers to become more comfortable with the researcher, each other, the recording booth and head-mounted microphones. The interviews are recorded for record-keeping purposes and potential future analyses of factors such as speaking style or task effects.

Map task

The *Map Task* is one of the find-the-difference list tasks intended to elicit stance-neutral first-mentions. Speakers are seated across from each other and given "maps" of imaginary superstores (provided in Appendix B). About 50 household items are listed in three columns representing aisles in a store. The two maps have the same items arranged in different orders; the task is to discuss all the items to determine how the arrangements differ without looking

Table 2.1: Map Task snippet

A:	My clothing items are at the bottom of th- of the third column. .. So I have things like jackets, shoelaces, socks, vests, coats, sweaters, boots, hats. [...] Boots, hats, backpacks, um - .. Although, backpacks, I would put that in with the camping supplies.

Table 2.2: Inventory Task snippet

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

at both maps. This task mostly consists of neutral exchanges of information, sometimes with comments on the logic of the arrangement. Table 2.1 provides an excerpt of a dyad completing this task.

Inventory task

The *Inventory Task* is a collaborative decision-making task designed to elicit low levels of involvement and weak stances. Speakers stand facing a felt-covered wall and are given a box of about 50 Velcro-backed cards that can be stuck to the felt. The cards are printed with the names of household items, and about 15 additional cards are already placed on the wall, which represents a store inventory map. Speakers are told to imagine that they are co-managers of a superstore in charge of arranging new inventory. Their job is to discuss each item in the box and agree on where to place it; once it is on the wall, it cannot be moved. This task generally involves polite solicitation and acceptance of suggestions; Table 2.2 provides an example exchange.

Table 2.3: Survival Task snippet

B:	Eighteen liters of water. That’s a lot of water. .. Just based on
A:	Yeah.
B:	the w- the weight. .. I mean, I - I took some fifty-mile hikes when I was in Boy Scouts. I know that .. the first thing you think about is how much does it weigh?
A:	Oh.
B:	Do you really wanna carry this - this stuff?
A:	Well .. we’re in a r- .. We’re in a raft,
B:	Okay. [...]
A:	So we can put it in the raft at first -
B:	That’s true.

Survival task

The *Survival Task* is a collaborative decision-making task designed to elicit moderate involvement and stances. Speakers are seated in front of the computer screen which explains the following scenario[2] (modeled after [82]): they are on a sinking ship near shore in sub-zero winter weather. They have a life raft, and the nearest town is 20 miles away. They have salvaged some items but cannot carry them all, so they must discuss each item and decide whether to take or leave it based on its usefulness for their survival. The items are the same as those used in the Map and Inventory tasks but with varying quantities (e.g., 5 socks, 1 coat). No constraints are placed on the number or types of items they can take. This task includes more negotiation and reasoning than previous tasks, sometimes including personal knowledge or experience to lend credibility, as in the except in Table 2.3.

[2] See [93] for descriptions of survival-scenario team-building exercises.

Table 2.4: Category Task snippet

B:	And I have - pothole maintenance is under infrastructure.
A:	That makes sense.

Category task

The *Category Task* is the find-the-difference list task intended to elicit stance-neutral first-mentions of the set of items to be used in the Budget Task. Procedures are the same as in the Map Task, but speakers are instructed to imagine that they are on a county budget committee, and their lists are the recommendations of two independent assessors tasked with identifying services or expenses that could be cut from various departments. Again, there are about 50 items on each list, but they are grouped into differing categories (e.g., transportation, education, public health), and speakers must find the differences without looking at both lists together. This task includes neutral exchanges of information, sometimes with added comments on item categorization or the importance of funding (or not) a service, as seen in the example in Table 2.4.

Budget task

The *Budget Task* is a collaborative decision-making task designed to elicit high levels of involvement and strong stances. Speakers are seated at a computer screen and told to imagine that they are on a county budget committee in charge of making cuts to four departments. About 50 services and expenses are divided among the four departments on the screen. Their job is to discuss each item and decide whether to fund or cut it; the only limitation is that they must cut the same number of items from each department. This task involves more elaborate negotiation, which may include citing personal knowledge or experience as support for stances. An example of this appears in the excerpt in Table 2.5.

Table 2.5: Budget Task snippet

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.

2.3 Transcription and annotation

After recordings are complete, the audio file(s) are copied and then cut in Praat into a separate stereo file for each task. Demographic interviews and free conversation between tasks are also saved separately for potential future use, e.g., in studies of speaking style. A Praat TextGrid is created for each task during the transcription and audio-alignment stages described in Section 2.3.1 below. New tiers are added to the TextGrid for annotations: stance strength and polarity, detailed in Section 2.3.2, and stance-act type, in Section 2.3.3. Transcription and annotation are prioritized for the Inventory and Budget tasks, the stance-dense activities designed to elicit the lowest and highest levels of involvement, so that the effects of such differences can be seen more clearly. A portion of the other tasks have been processed but are not analyzed in the current studies.

2.3.1 Transcription and forced-alignment

Tasks are manually transcribed at the utterance level in Praat [10] following a simplified version of the ICSI Meeting Corpus transcription guidelines [69]. Each speaker is transcribed in a separate interval tier of a TextGrid. Stretches of speech are demarked when surrounded by at least 500 ms of silence, and every word of the resulting 'spurt' is transcribed orthographically using conventional American spelling, with the addition of common shortenings (cuz, kay, etc.), phonological contractions (gonna, wanna, hafta, kinda, etc.), discourse markers (uh-oh, mm-hm, etc.), and vocalizations with recognized meanings (e.g., psst, shh, meh (verbal shrug), psh (verbal scoff)). Numbers and symbols are spelled out, and pronounced letters are transcribed as capitalized with a following underscore (e.g., A_B_C_). Pauses shorter than 500 ms are marked within an utterance with two periods. Filled pauses are transcribed as “uh” or “um,” with the latter indicating audible nasality. Disfluencies are marked with a short dash, without a space for truncated words (e.g., categ-) or following a space for uncompleted thoughts (e.g., I thought -), which may end an utterance or precede a repetition or restart (e.g., I don’t - I’m not - I’m not sure.). A small, finite set of vocalizations is transcribed with tags (e.g., {VOC laugh}, {VOC cough}), and notable voice qualities or unusual pronunciations are marked with a following descriptive tag (e.g., {QUAL laughing}). Utterances are transcribed using conventional capitalization and a limited set of punctuation marks, e.g., period to end a complete statement, question mark to end a syntactic question, commas to separate lists (no colons, semi-colons, or quotation marks are used). For the full list of conventions, see the transcription guidelines in Appendix C.

Completed manual transcriptions are automatically force-aligned using the Penn Phonetics Lab Forced Aligner (P2FA [108]), which demarks word and phone boundaries in separate interval tiers for each speaker in the Praat TextGrids. Transcribed words not already in the pronouncing dictionary provided with P2FA (CMUdict [102]) (place names, truncations, vocalizations, etc.) are added as needed. The dictionary uses ARPAbet [87], a transcription system which assigns each consonant a one- or two-letter code and each vowel two letters plus a digit for lexical stress (1 for primary, 2 secondary, 0 unstressed). However, International Phonetic Alphabet (IPA) symbols are used here for clarity of presentation[3].

2.3.2 Stance strength and polarity annotation

After orthographic transcription and forced-alignment, the tasks are manually annotated at a coarse, inter-pausal level 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) is marked

[3] Correspondences between IPA and ARPAbet symbols can be found at `http://www.speech.cs.cmu.edu/cgi-bin/cmudict` and many other websites.

with one of the stance presence/strength labels shown in Table 2.6. Spurts with a discernible stance strength (label 1, 2, or 3) are also labeled for polarity, as shown in Table 2.7. As a result, each spurt is marked with one of 14 possible strength-polarity label combinations. The full annotation guidelines appear in Appendix D.

Table 2.6: Stance strength levels

Label	Description and examples
0	No stance: list reading, backchannels, fact-exchange, 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 no-stance content, e.g., “Ooh, buckets!” “I don’t know what that means.”

Table 2.7: 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.

Both textual content and prosody are taken into account when determining labels, as prosody can be used to enhance or even reverse the meaning of text alone. Because strength is relative, the scheme is applied on a per-speaker basis. Before labeling, annotators listen 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 pitch and intensity ranges, small deviations are more meaningful than for the most energetic speakers, whose modulations must be more extreme to indicate differences in stance. Annotators listen to the audio in Praat while labeling one speaker's transcription on a new interval tier in the TextGrid, and then listen again while labeling the other's in a separate tier.

The scheme was verified for its usability with triple blind 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 ($p = 0$). This level of agreement was deemed sufficient to allow less overlap in annotation in favor of an overall faster procedure. After a task is labeled by one annotator, a second reviews and verifies or corrects each label while listening to the audio. Asterisks (*) are used to indicate uncertainty, with the second annotator providing a second opinion as needed. If the second annotator remains uncertain about a label, a third annotator serves as a tie-breaker. In the 20-dyad sample used for acoustic analysis in Experiments 2-3 (Chapters 5-6), 5.4% of spurts are 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 yields very high inter-rater agreement. Weighted Cohen's kappas with equidistant penalties are 0.87 for stance strength labels and 0.93 for for polarity labels ($p = 0$), with the unweighted kappa for combined labels at 0.88 ($p = 0$).

The approach of labeling spurts rather than a more structurally-based linguistic unit, such as clauses or sentences, allows for a holistic view that is relatively quick to annotate. However, for a more nuanced taxonomy of stance components and types, which may display substantially different acoustic cues, annotation must occur at a more detailed level.

2.3.3 Stance type annotation

To begin separating types of stance moves at a more fine-grained level, annotators label only words and phrases which perform 'stance acts,' or dialog acts involving stance-taking (cf. e.g., [14, 23]). Annotators listen to both speakers' channels while annotating first one speaker and then the other, in separate interval tiers in a Praat TextGrid. While spurts are used as a unit of convenience for stance strength and polarity annotation (Section 2.3.2), stance act boundaries are determined by the annotators, and acts may divide or span multiple spurts. Both lexical and prosodic information is considered when choosing the lexical makeup of a stance act, based on how it performs the functions shown in Table 2.8 within the discourse context. (See Appendix E for the full annotation guidelines.) This stance-act type annotation scheme draws on a range of content- and discourse analytic literature with a variety of stance-related concepts and classifications (cf. [52]), as described below.

Some of the most overt types of stance-taking are included in the *opinion-offering* category (o): evaluation and evaluative description [19, 20, 23, 50, 57, 58], appraisal, judgment, appreciation, affect/affective stance [65, 72], assessment [47, 72], subjectivity, intersubjectivity, positioning, alignment [19, 20, 46, 54, 57], attitude/attitudinal stance [19, 65, 72], recommendation, persuasion, modality, modulation [19, 23, 45], and prediction [19, 50].

In the *convincing/credibility* category (c), speakers engage 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 [7, 19, 23, 45, 47, 50, 57]. *Hedging, softening, or hesitation* to offer a stance (f) may 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 [7, 19, 23, 45, 50, 57, 65]. It could also be used for interpersonal stance, e.g., to show deference to another's preferences or authority [11, 50].

In *soliciting* another's stance (s), speakers engage in both knowledge exchange [23] and interpersonal stance-taking, (also called performative positioning [46]), which involves ne-

Table 2.8: 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 still feels “stancey”)
b	Backchannels (e.g., “Mm-hm, yeah”)
0	No-stance (unlabeled for stance type, e.g., factual questions and answers)

gotiating 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 [20, 47, 50, 54]. Both *teamwork/rapport-building* and *encouragement/praise* (t, e) are interpersonal in nature [20, 47, 54, 57], 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 [57] in that they take stances in relation to previous stances of any type [19, 20, 23, 47, 72]. 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 [11, 20, 23, 50, 72].

The remaining categories allow for types of stance that are difficult to name (*expressive intonation, unclear* (i, x)) and those which normally carry little or no stance (*backchannels, no-stance* (b, 0)). Backchannels are separated due to their recognizable discourse function and previously-studied acoustic properties (cf. e.g., [6, 39, 95, 98]), which may serve as a useful basis of comparison against stronger stance types.

Some of the labels serve similar functions which are often more difficult to differentiate during annotation. A distinguishing feature between *agreement* and *opinion-offering* (a, o) is whether the utterance takes a new stance (o) or merely shows acceptance/approval of an existing one (a). Similarly, lexically positive *backchannels* (b) like 'yeah, right, okay' can be difficult to distinguish from *agreement/acceptance* (a); here the rule of thumb is whether the speaker takes (or attempts to take) the floor (a). (The new turn may continue after the agreement, or if the agreement is the entire turn, the other speaker often begins a new turn in response, whereas backchannels generally occur during another speaker's turn.) While *reluctance* and *hedging* (r, f) can sound similar, *reluctance* usually occurs in response to another's stance to soften or avoid rejection, while *hedging* attempts 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 are 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) vs. sarcasm or commiseration (t), which on the surface may appear negative but which serve to build solidarity (i.e., "At least we're in the same boat"). Finally, categories for general and intonationally-carried "stanciness" (x, i) are left underspecified to allow for additional classifications that may emerge in future analyses.

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) would be labeled (os). Uncertainty in an initial label choice is indicated with an added asterisk (*), which is removed after all annotations are verified or modified by two additional annotators. In the 20-dyad sample used for acoustic analysis below and in Experiments 2-3 (Chapters 5-6), 5% of acts are marked with uncertainty by a first annotator, and only 1% by a second. Labels receiving greater than 5% initial uncertainty include: *reluctance, disagreement, opinion with reasons, softened opinion, intonation*, and *unclear* (r, d, co, fo, i, x). Finally, a pound symbol (#) is added to any stance act label in which the automatic forced-alignment deviates significantly from the audio signal. These poor alignments make up a small proportion of the recordings (4.3% of acts in the 20-dyad sample), and so they are removed from current analysis but could be corrected should future analyses require them.

2.4 Structural and acoustic characteristics

The result of the above procedures is the large audio corpus containing 68 speakers in over 40 hours of conversational interaction. This section describes the corpus size and makeup in terms of speakers, tasks, and general acoustic properties.

2.4.1 Speakers

As mentioned above, all speakers in the ATAROS corpus are native English-speakers age 18-75 who grew up in the Pacific Northwest (Washington, Oregon, Idaho). The majority, 54 speakers, grew up primarily in Western Washington, mostly in the Seattle metropolitan area;

Table 2.9: Dyads by age and sex

		Dyads by sex			
Group	Ages	FF	MM	MF	*Sums*
Younger	(18-32)	5	2	9	16
Middle	(38-49)	1	2	3	6
Older	(60-75)	7	1	4	12
Totals		13	5	16	34

Table 2.10: Speakers by age and sex

		Speaker sex		
Group	Ages	F	M	*Sums*
Younger	(18-32)	19	13	32
Middle	(38-49)	5	7	12
Older	(60-75)	18	6	24
Totals		42	26	68

7 are from the Portland, OR/Vancouver, WA area, 3 from areas farther south in Western Oregon, and 4 from Central or Eastern Washington and Northern Idaho. They are paired to form dyads of strangers matched roughly by age and crossed or matched by sex. Table 2.9 shows the distribution of dyad compositions by age and sex. Of the 34 dyads recorded, 13 are female-female, 5 male-male, and 16 mixed-gender, yielding a total of 42 females and 26 males (68 total speakers). Table 2.10 shows the same data but in terms of speaker sex rather than dyad composition. About half (47%) the speakers are under age 35 (19 females, 13 males), a third (35%) over age 60 (18 females, 6 males), and less than a fifth (18%) age 35-60 (5 females, 7 males). The exact age ranges of speakers in these rough age groups are shown in the tables. In the future, if more speakers age 40-60 are recorded, the current 'middle' age group may be split, and/or the boundary between 'young' and 'middle' may shift.

2.4.2 Corpus size

Each task takes about 13 minutes to complete on average (range: 3-42 minutes), yielding an average of 64 minutes of collaborative interaction per dyad (range: 35-154 minutes). In total, the corpus contains 36.4 hours of collaborative interaction, plus 5 hours of demographic interviews and miscellaneous conversation for a total of 41.4 hours of dyadic speech.

To date, all Inventory and Budget tasks have been transcribed and time-aligned, except for one dyad which has so far been excluded from analysis due to a high frequency of unusual mannerisms (cf. Section 2.2.1). Stance strength, polarity, and type annotation is complete for 26 dyads' Inventory and Budget tasks. Transcription, alignment, and stance type annotation of the other collaborative tasks is nearly complete for 8 dyads and will continue in the future.

2.4.3 Task differences

Given that the tasks are intended to encourage differing levels of involvement and stance-taking, some initial task validation is presented here in order to explore any systematic differences in speaking style between tasks.

As mentioned above, dyads spend similar amounts of time on each task, about 13 minutes on average, but the range of task lengths can be quite large, with the Inventory task showing the most consistency. Table 2.11 shows the mean, standard deviation, and range of lengths (in minutes) for each task, with the stance-neutral tasks (Map, Category) separated from the stance-dense tasks, arranged in order of increasing involvement.

Before beginning acoustic analysis, members of the ATAROS research team applied measures that can be directly extracted from the time-aligned transcripts to the first 12 transcribed dyads' Inventory and Budget tasks, the two stance-rich tasks designed to elicit the

Table 2.11: Task length statistics

	Length (min)			
Task	Mean	St.Dev.	Min.	Max.
Map	11.1	6.1	5.0	31.7
Category	12.3	6.1	6.6	34.8
Inventory	12.5	4.1	6.7	23.0
Survival	14.7	7.9	7.5	42.4
Budget	13.6	7.8	2.6	39.9

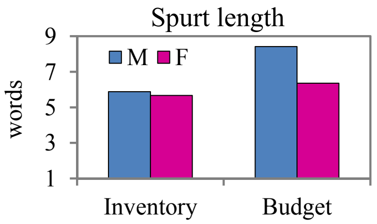


Figure 2.1: Spurt length by task and sex: 12-dyad sample

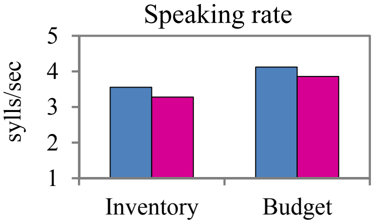


Figure 2.2: Speaking rate by task and sex: 12-dyad sample

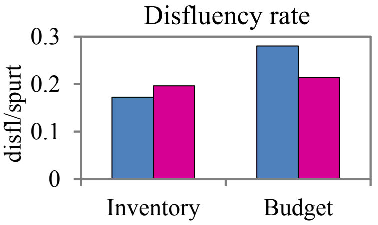


Figure 2.3: Disfluency rate by task and sex: 12-dyad sample

lowest and highest levels of involvement, and those prioritized for further annotation and analysis. These dyads are evenly split by sex, with 3 female-female, 3 male-male, and 6 mixed. As reported in [28, 29], average task lengths are comparable between tasks, in terms of total time spent ($\approx$ 11.5 minutes), number of transcribed words ($\approx$ 1750), and number of turns between speakers ($\approx$ 275). As with the entire corpus (Table 2.11), total task length is less variable for the Inventory task (standard deviation 2.25 minutes compared to 5 minutes for the Budget task). Based on Wilcoxon signed-ranked tests, the Budget task shows significantly longer utterances (in mean number of words per 'spurt,' or stretch of speech between silences of at least 500 ms, $p < 0.001$, Figure 2.1) and significantly faster speaking rates (in vowels per second, vps, a proxy for syllables/sec, $p < 0.001$, Figure 2.2). Three types of disfluencies are easily extracted from the transcriptions: filled pauses ("uh, um") and truncated words are transcribed directly, and repetitions between speakers can be automatically detected using a model trained on the SWITCHBOARD corpus [37, 74, 109]. With filled pauses and truncated words counted together, the rate of such disfluencies per spurt is significantly higher in the Budget task ($p < 0.05$, Figure 2.3), as are repetition rates between speakers ($p < 0.01$). Interestingly, males appear to exhibit a larger difference between tasks, with both spurt lengths and disfluencies increasing by about a third in the Budget task compared to the Inventory task (Figures 2.1 and 2.3).

Similar measures are applied to the 20-dyad sample of Inventory and Budget tasks used

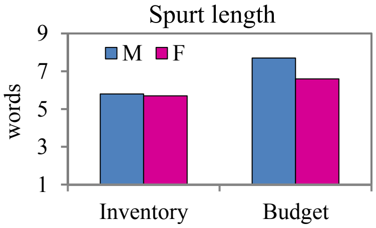


Figure 2.4: Spurt length by task and sex: 20-dyad sample

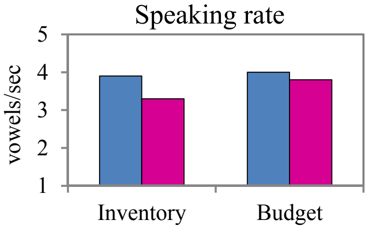


Figure 2.5: Speaking rate by task and sex: 20-dyad sample

for acoustic analysis in Experiments 2-3 (Chapters 5-6). Mean speaking rate in vowels per second (vps) and mean spurt length in number of words are calculated for each speaker within each task, and Wilcoxon signed-rank tests are used to examine task effects with all speakers pooled and within each sex. Overall, the two tasks are similar in total duration (Table 2.11), and in the total number of spurts uttered per speaker (≈ 148), but the Budget task again exhibits significantly longer spurts (mean 7 words, compared to 5.7 in the Inventory task, $p < 0.001$, Figure 2.4) and faster speaking rates (Figure 2.5). As in the 12-dyad sample, the task effect holds within each sex ($p < 0.01$), and Budget spurt lengths increase more for males, but speaker sex has no effect on spurt length, either alone or within each task (Figure 2.4). (These effects also hold for mean spurt duration in seconds.) In contrast to the 12-dyad sample, speaking rate shows no task effect for men, who again exhibit slightly faster rates in both tasks, but rather, the overall task effect is driven by women, who speak more slowly in the Inventory task (mean 3.3 vps, compared to 3.8 in the Budget task, $p < 0.001$, Figure 2.5). This pattern is driven by women's mean unstressed vowel durations, which are significantly shorter in the Budget task than in the Inventory task ($p < 0.001$), while no task effects are found for stressed vowel duration. In other words, while the average speaking rate for men does not vary between tasks, women speak more slowly in the Inventory task but match men's rates in the Budget task via greater unstressed vowel reduction.

These findings – longer utterances, faster speech, increased disfluencies – are consistent with higher levels of involvement, as intended by the task design. With a generally similar amount of speech obtained in each task, and considering these task-related style differences, task effects are explored within each measure reported here, although their impact on vowel-level acoustic measures is normally small or absent, allowing for data from both tasks to be combined.

2.4.4 Stance labels

Given the annotation protocols for stance strength and polarity (Section 2.3.2), uneven distributions across levels are expected. Table 2.12 shows the numbers of spurts with each strength/polarity label in the 20-dyad sample of Inventory and Budget tasks used for acoustic analysis in Experiments 2-3 (Chapters 5-6). Of the nearly 11,500 labeled spurts, about 47% have weak stance strength, 22% moderate, 1% strong, 24% none, with the remaining 6% unclear. Spurts with clear stance strength (weak, moderate, strong) are also labeled for polarity; overall, about 37% of these spurts receive positive labels and 7% negative, leaving 56% with neither positive nor negative polarity.

Because stance acts are delimited independent of spurt boundaries (Section 2.3.3), they

Table 2.12: Spurts by stance strength and polarity

		Polarity					
		Positive	Neutral	Negative	Unclear	NA	
Strength		(+)		(−)	(x)		*Sums*
None	(0)	-	-	-	-	2812	2812
Weak	(1)	2565	2760	76	4	-	5405
Moderate	(2)	422	1721	419	9	-	2571
Strong	(3)	6	38	22	2	-	68
Unclear	(x)	-	-	-	-	633	633
Totals		2993	4519	517	15	3445	11489

differ in structure from spurts. On average, stance acts in the 20-dyad sample are 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 is unaffected, at about 3 words per second.) As with spurts, stance acts are longer on average in the Budget task (mean 4.4 words, compared to 3.5 in the Inventory task). These patterns holds for both sexes.

Stance act type labels are also unevenly distributed. Of the nearly 12,400 acts in the 20-dyad sample, about half are divided between *agreement* and *opinion-offers* (a, o), with another quarter spread over *convincing, offer+solicit, soliciting*, and *hedging/hesitation* (c, os, s, f). These distributions are shown in Table 2.13, separated from the remaining six types with greater than 100 acts, which each constitute 1%-2% of labeled acts. (In addition, there

Table 2.13: Stance acts by type: Types with > 100 acts

Stance type		N acts	Mean length (words)
a	agreement	3292	1.9
o	opinion	3000	5.9
c	convincing	1564	8.7
os	offer+solicit	703	5.3
s	soliciting	393	3.7
f	hedging	345	3.0
co	opinion+cred	267	9.0
b	backchannel	193	1.1
x	unclear	188	2.7
r	reluctance	184	2.2
t	rapport	158	4.9
ot	opin+rapport	137	7.0

are over 3400 stretches of speech unlabeled for stance; these are not considered stance acts, but their vowels are included in acoustic analyses for comparison purposes.) As is clear from the table, stance act types vary substantially in length. *Backchannels* (b) are generally one-word acts, and markers of *agreement, reluctance, hedging*, and *opinion-soliciting* (a, r, f, s) are about 2-4 words long. Other types are much longer, with acts involving *convincing/reasoning* (c, co) taking about 9 words, and others involving *opinion-offering* or *rapport* (o, os, ot, t) 5-7. A one-way ANOVA assuming unequal variance applied to these 12 most frequent stance act types shows that the relationship is significant ($F[11, 1290] = 683.8, p < 0.001$). (This also holds for stance act duration measured in seconds.)

2.4.5 Acoustic properties

To begin characterizing the corpus on acoustic measures, vowels are examined from the sample of 20 dyads' Inventory and Budget tasks described in Experiments 2-3 (Chapters 5-6). As detailed in Sections 6.1-6.2, the sample contains over 89,000 vowels with automatically-extracted acoustic measures, normalized within-speaker (using z-scores for pitch and intensity, z-scores within vowel quality for duration, and the vowel-extrinsic 'Nearey 2' method [71] for formants). This section explores the effects of several known factors on prosodic measures (pitch, intensity, and duration) and vowel spaces. These factors include: lexical stress, as assigned by the pronouncing dictionary (CMUdict [102]) during time-alignment of the transcripts; grammatical function of the words containing the vowels (content/function), assigned with reference to a list of English function words (prepositions, pronouns, auxiliaries, determiners, conjunctions, etc.) modified from [68]; spurt- and sentence location of the words (initial, medial, final, or single for single-word spurts/sentences); vowel quality; speaker sex; and task (Inventory/Budget).

Prosodic features

Several of the above factors have effects on vowel pitch. Separate one-way ANOVAs (assuming unequal variance) applied to pitch at midpoint for all measured vowels show significant effects for: lexical stress ($F[2, 4578] = 38.0, p < 0.001$), spurt location ($F[3, 7013] = 301.9, p < 0.001$), sentence location ($F[3, 13511] = 386.1, p < 0.001$), and vowel quality ($F[14, 7349] = 50.1, p < 0.001$), with pitch level roughly corresponding to vowel height, as expected (cf. summary in [73]). Welch's two-sample t-tests show that pitch is significantly higher in vowels with primary lexical stress, and it declines over the duration of utterances and sentences ($p < 0.01$). Pitch is not significantly affected by task or content/function-word status, and sex differences are neutralized via within-speaker normalization. (In raw Hz, mean pitch for females is about 190 Hz (median 189, range 148-221), and for males about 115 Hz (median 114, range 93-148).)

For intensity at vowel midpoint, separate one-way ANOVAs (assuming unequal variance) show significant effects for: lexical stress ($F[2, 7386] = 1024.0, p < 0.001$), content vs. function words ($F[1, 85930] = 23.2, p < 0.001$), spurt location ($F[3, 11745] = 482.1, p < 0.001$), and sentence location ($F[3, 22311] = 413.2, p < 0.001$). Welch's two-sample t-tests show that intensity is significantly higher ($p < 0.001$) in lexically stressed vowels, and content words; it declines over the duration of utterances, and drops at the ends of sentences. Intensity is not significantly affected by task or sex.

For vowel duration, separate one-way ANOVAs (assuming unequal variance) show significant effects for: lexical stress ($F[2, 7779] = 739.9, p < 0.001$), content vs. function words ($F[1, 82078] = 39.7, p < 0.001$), spurt location ($F[3, 11218] = 1156.3, p < 0.001$), and sentence location ($F[3, 20766] = 2013.4, p < 0.001$). These factors are commonly known to have durational effects, and in the sample, Welch's two-sample t-tests show that all behave as expected: durations are significantly longer ($p < 0.001$) in lexically stressed vowels, content words, at the ends of sentences and utterances, and even longer in single-word utterances. Durations also differ significantly between tasks, with longer durations in the Inventory task ($F[1, 84650] = 30.7, p < 0.001$), consistent with the faster speaking rates reported for the Budget task in Section 2.4.5. Speaker sex does not have a significant effect on vowel duration, and vowel qualities are not comparable, as normalization is done within vowel quality to reduce effects of intrinsic durational differences.

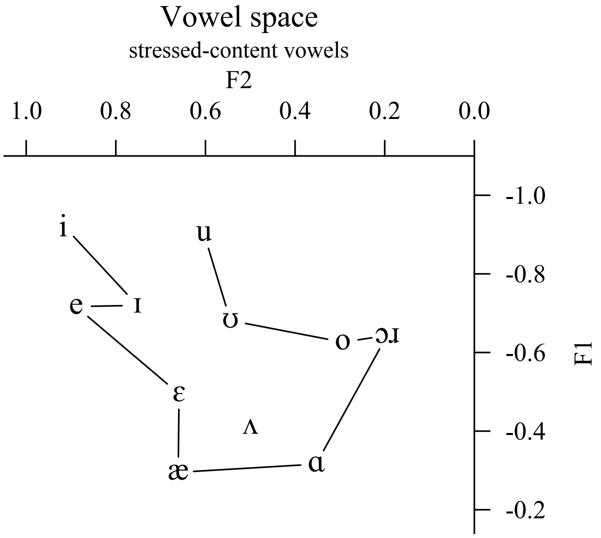


Figure 2.6: Vowel space: stressed-content vowels (Nearey2-normalized)

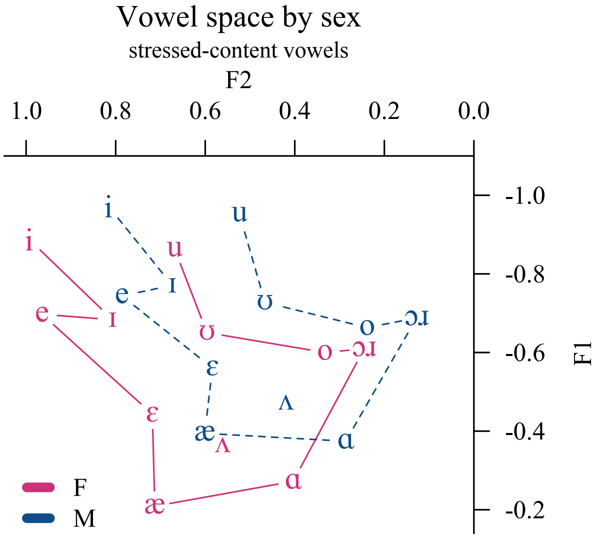


Figure 2.7: Vowel space by sex: stressed-content vowels (Nearey2-normalized)

Vowel spaces

As elaborated in Section 6.2, formant measures are automatically extracted from vowels marked in the time-aligned transcripts and then normalized within speaker using the vowel-extrinsic 'Nearey 2' method [71]. To reduce outliers caused by tracking or alignment errors, about 5% of the most extreme (highest/lowest) measurements are excluded from each vowel quality said by each speaker. This section explores the vowel space with regards to the same factors presented for prosodic measures above (lexical stress, word function, spurt/sentence location, speaker sex, task).

The general arrangement of the vowel space is shown in Figure 2.6, which plots mean normalized F1xF2 measures taken at midpoint of the automatically-measured stressed vowels in content words said by all 40 speakers in the sample reported in Experiments 2-3 (Chapters 5-6). The vowel positions are consistent with those reported in other work on Pacific Northwest English (PNWE) (e.g., [3, 18, 27, 51, 83, 85, 97, 99, 107]). As one of the defining features of Western American speech (cf. e.g., [17, 38, 60]), the low-back merger of /a/

and /ɔ/ is present in PNWE, except before /ɹ/, where /ɔɹ/ merges with /oɹ/ but not /aɹ/ [24, 99]. This is confirmed for the sample by splitting stressed vowels with /ɔ/ assignments in the pronouncing dictionary (CMUdict [102]) used by the forced-aligner (P2FA [108]) by following word-internal phone (/ɹ/ or other). Pre-rhotic /ɔ/, labeled /ɔɹ/ on the plots, remains separate in a mid-low back position near /o/ while non-pre-rhotic /ɔ/ is merged with /a/ in the low back corner and is therefore collapsed with /a/ and labeled /ɑ/ in all vowel plots presented here. Also common in Western dialects, /u/ and /ʊ/ are fronted in the sample, but without the accompanying fronting of /o/ found in California and Oregon [3, 18, 21, 44, 55, 63, 67, 97]. The back position for /o/ fits the patterns found in Seattle [27, 99] and other areas of Washington State [85] but not the fronted position found in Portland, Oregon (another prominent Northwest city) [3, 18, 97], which is unsurprising given that most speakers in the sample are from the Seattle area (79%, compared to 10% from the Portland area). Similarly, /æ/ in the sample[4] is low-front as in Washington [27, 83, 85, 99, 107], rather than backed, as in the neighboring dialect regions of Oregon, California, and Canada [3, 9, 16, 18, 21, 44, 55]. The other front lax vowels /ɛ, ɪ/ pattern with both Seattle and Portland in showing little lowering or backing [3, 85, 99] when compared to the shifts seen in California, Canada, and the Northern Cities [9, 16, 21, 44, 55, 59, 60]. Finally, /i, e/ are high in the system, and /ʌ/ is central and fairly low, as observed in other work on the West and PNWE [17, 60, 85, 101, 99].

Several known factors affect vowel spaces as expected: males' spaces are shifted and slightly compressed compared to females' (Figure 2.7), lexically unstressed vowels are more reduced and more variable than stressed (Figure 2.8), as are vowels in function words compared to those in content words (Figure 2.9). Due to the reduction and wider variation of function words and unstressed vowels, most measures and plots presented here involve only stressed vowels in content words ('stressed-content vowels'). Because the shapes of male and female spaces are similar, data from both sexes are combined, except when examining soci-

[4] Note that /æ/ is reported here in all phonetic environments combined; extensive work in many dialect regions predicts possible departures from this mean in a variety of environments.

olinguistic gender effects. Similarly, vowel measurements are not separated by task, spurt- or sentence location because these factors do not show an effect on vowel space arrangements.

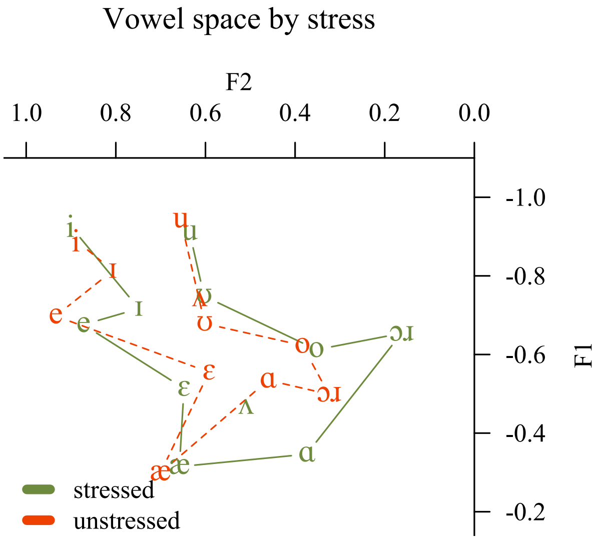


Figure 2.8: Vowel space by lexical stress (Nearey2-normalized)

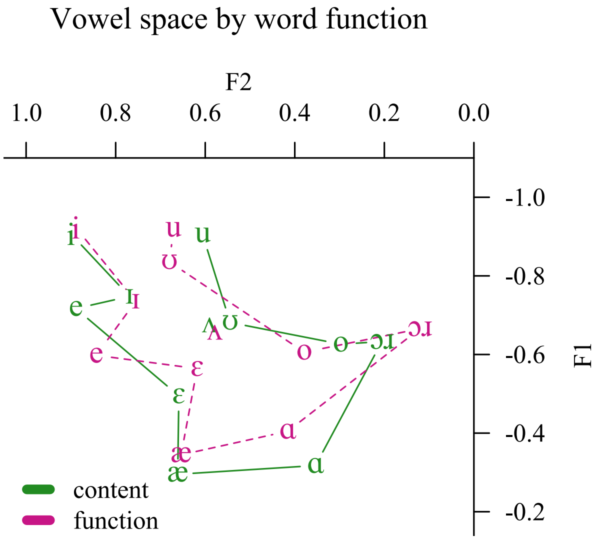


Figure 2.9: Vowel space by word function (Nearey2-normalized)

Chapter 3
PREDICTIONS

Many layers of meaning are conveyed in natural speech, beyond the words and their syntax. One such layer is stance, or the expression of an attitude toward an object, claim, or person relevant within the discussion context [8, 20]. Previous work has found that variation in aspects of pronunciation associated with prosody (e.g., vowel duration, speech rate, pitch excursion) reliably differentiate stance-expressing phrases from neutral utterances in unscripted speech [25, 26]. The work presented here builds on these early findings with a more fine-grained approach to the investigation of stance-taking and its relationship to acoustic variation in spontaneous conversation. It takes up the argument that since stance presence is signaled acoustically, components or features of stance may differ acoustically as well. Two holistic features are examined, stance strength and polarity, as well as categories of more specific stance act types.

The central prediction of this work is that **stance type, strength, and polarity are signaled by changes in the acoustic signal**. Three studies are conducted to test this prediction using speech taken from the ATAROS corpus (Chapter 2), and acoustic measures associated with prosodic and vowel space features. Experiment 1 (Chapter 4) examines the pitch and intensity contours of a small sample of instances of the word ‘yeah’ that contribute to negative stances. Experiment 2 (Chapter 5) investigates over 2200 ‘yeahs’ for prosodic cues to stance type, strength, and polarity. Experiment 3 (Chapter 6) expands the investigation of prosodic cues to the stressed vowels in all content words spoken by 40 speakers in the Inventory and Budget tasks.

Chapter 4

EXPERIMENT 1: THE PROSODY OF NEGATIVE 'YEAH'

Stance – attitudes and opinions – can be expressed with a combination of lexical and acoustic features which often work in concert toward particular meanings, for example in conveying a positive or negative message. Normally, 'yeah' has positive polarity; it is used to agree, affirm, accept, etc. However, with a change in prosody, 'yeah' can also convey a negative stance, e.g., in expressing polite disagreement or echoing another's negative sentiment. Since its lexical content is by default positive, negative meanings must be carried in the speech signal. This study[1] investigates acoustic-prosodic features of such 'negative yeahs' by examining the pitch and intensity contours that distinguish four subtypes of negative 'yeah' as identified through content analysis. It serves as a pilot for the larger study of the prosody of 'yeah' reported in Chapter 5.

4.1 Corpus sample and stance functions

Building on the preliminary coarse-grained analysis of stance acts in the ATAROS corpus described in Chapter 2, the cue word 'yeah' was identified as a good candidate for further investigation; it occurs frequently in the corpus and is associated with a variety of stance acts, ranging from discourse-functional backchannels (typically with no or weak positive stance) to emphatic agreement (strong, positive stance) [40, 41, 49]. In the course of extracting 'yeah' tokens, it was noticed that while most occur in positive or neutral utterances, as expected, a few appear to contribute to negative stances, contrary to their presumed positive lexical polarity. They are examined here as an examples of how acoustic-prosodic features can

[1] Portions of this work were presented to the Linguistics Society of America (LSA) and published in the 2015 *LSA Annual Meeting Extended Abstracts* [33].

change the meaning of identical lexical material, even to the point of changing its polarity.

The dataset for this study is comprised of natural speech from the first 23 dyads (46 speakers) in the ATAROS corpus to be transcribed and annotated for stance strength and polarity as described in Chapter 2. The sample consists of 'yeahs' uttered during two of the collaborative tasks: the Inventory task, in which dyads arrange household items to make a map of an imaginary superstore, and the Budget task, in which they choose services to cut from an imaginary county budget. This sample yields 8.7 hours of conversation and a total of 2870 'yeahs' (54% said by males, 46% by females). The majority of 'yeahs' (68%) occur in positive-marked utterances, indicating agreement, encouragement, etc. About 30% occur in neutral or non-stance utterances (backchannels, acknowledgments, etc.). Only 61 'yeahs' (2%) occur in negative-marked utterances, which were examined further to identify more specific discourse functions of the 'yeahs' using content analysis similar to that conducted for the stance type annotation described in Section 2.3.3. After excluding positive and unclear uses, only 46 'yeahs' said by 24 speakers clearly contribute to the negative stances of their utterances. All but three of these cluster into the four common categories that emerged from this analysis, labeled as follows:

(a) *yeah but* (N=12 utterances): 'yeah' is quickly followed by an explanation against a preceding stance

(b) *reluctance* (N=13): 'yeah' indicates reluctance to accept or agree with a previous stance

(c) *tough problem* (N=12): 'yeah' contributes to an expression of shared difficulty (e.g., "Yeah shoot, this is a tough problem.")

(d) *that's bad* (N=6): 'yeah' states agreement with a negative assessment without the empathy implied in the *tough problem* category (e.g., "Yeah you're right, that's bad.")

4.2 Analysis

For all 2870 'yeahs' in the corpus sample, intensity and pitch were measured via a script in Praat [10] at every decile of word duration and then z-score normalized speaker-internally to enable cross-speaker comparison. With all 'yeahs' examined together, both pitch and in-

Table 4.1: Pitch and intensity contours of negative ‘yeahs’ by stance function

Contours	Flatter intensity	Domed intensity
Flat pitch	*tough problem* (N=12)	*that’s bad* (N=6)
Contour pitch	*reluctance* (N=13)	*yeah but* (N=12)

tensity increase with stance strength, and negative ‘yeahs’ display slightly higher pitch and intensity than positive/neutral ‘yeahs.’ Looking at only the negative ‘yeahs,’ the four categories listed above are distinguished by an interaction of pitch and intensity patterns over the course of the word, as summarized in Table 4.1 and illustrated in the smoothing-spline ANOVA plots in Figures 4.1-4.2, which resemble aggregate pitch and intensity traces on a spectrogram by displaying splines connecting mean values at each measurement point, surrounded by shading representing 95% confidence intervals around the means (cf. [42, 100]). ‘Yeahs’ spoken in utterances in the *tough problem* and *that’s bad* categories (abbreviated *problem* and *bad* in the figures) have lower, flat pitch, while ‘yeahs’ in *reluctant* utterances have a high dipping contour and those in the *yeah but* category (abbreviated *but* in the

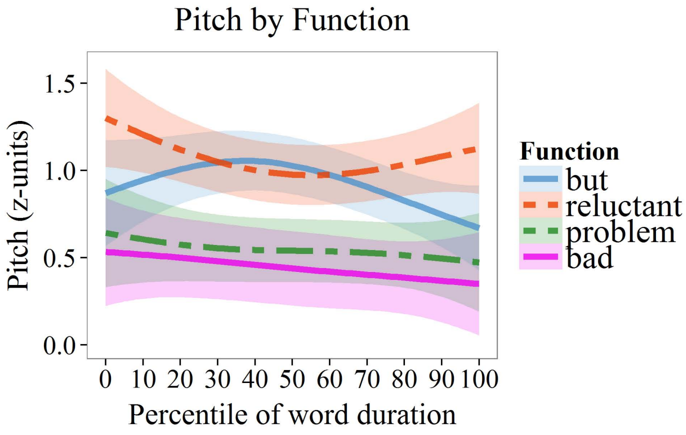


Figure 4.1: Pitch contours: negative ‘yeah’

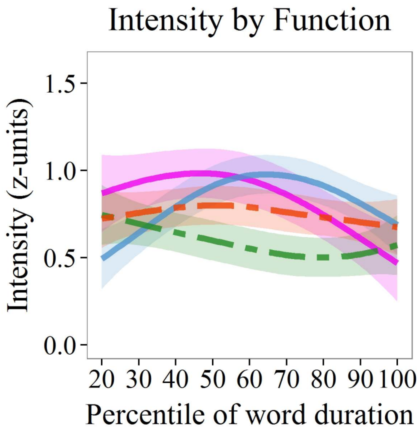


Figure 4.2: Intensity contours: negative ‘yeah’

figures) display a medium-high domed contour (Figure 4.1). Cross-cutting these groups, *reluctant* and *tough problem* ‘yeahs’ have lower, flatter intensity contours, while those in *yeah but* and *that’s bad* utterances have higher, domed contours (Figure 4.2[2]).

4.3 Discussion

These patterns show that fine-grained stance analysis can reveal word-level acoustic patterns that are not apparent in coarser approaches. With the small sample size, no claims can be made about whether the exact contour shapes or configurations apply to either ‘negative yeah’ or the described subcategories in general; rather, the key point is that qualitative methods, such as the content analysis and stance-annotation used here, can work in concert with combinations of acoustic measures to identify patterns in the speech signal that speakers use to convey – and understand – various subtle messages, whether propositional, social, attitudinal, emotional, etc.

Using similar approaches, future analysis of polarized lexical material will include the subcategorization of positive-marked ‘yeahs’ and comparison with negative words like ‘no’ (cf. [26]). Initial analysis suggests that without subcategorization, both of these groups’ pitch and intensity patterns occupy a range intermediate to that of the four categories of negative ‘yeahs.’ Figures 4.3-4.4 show the smoothing-spline ANOVA plots from Figures 4.1-4.2 with two new splines added: one for the 2824 ‘yeahs’ in the corpus sample that occur in positive or neutral stance-marked utterances (labeled *yeah+*), and one for the 246 instances of ‘no’ in the sample. The intermediate, relatively flat contours strongly resemble the shapes and locations of splines with all negative ‘yeahs’ combined, suggesting that positive/neutral ‘yeahs’ and ’nos’ may also reflect diverging subcategories that could be differentiated via more detailed stance type classification. This is the basis for the study of a larger sample of ‘yeahs,’ described in Chapter 5.

[2] First two deciles removed from intensity plots due to tracking errors and missing data.

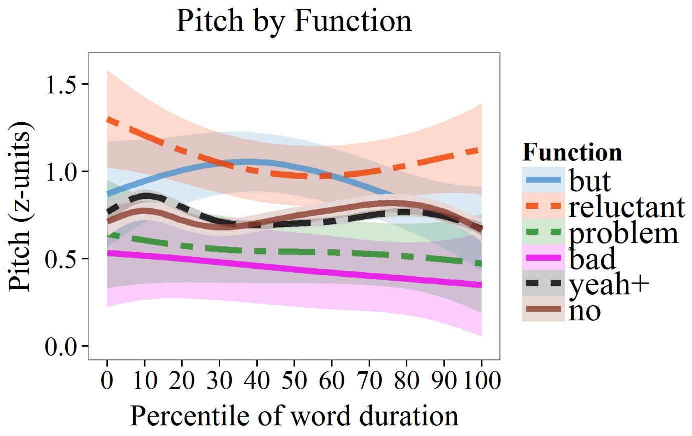


Figure 4.3: Pitch contours: negative 'yeah,' positive/neutral 'yeah,' 'no'

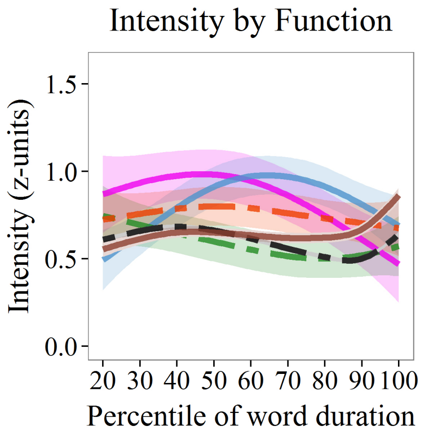


Figure 4.4: Intensity contours: negative 'yeah,' positive/neutral 'yeah,' 'no'

Chapter 5

EXPERIMENT 2: PROSODIC FEATURES OF STANCE IN 'YEAH'

This study[1] investigates prosodic characteristics of stance type, strength, and polarity in uses of the word 'yeah.' In a sample of 20 talker dyads engaged in two collaborative tasks, over 2300 'yeahs' fall into six common stance-act categories (Table 5.1). While *agreement*, usually with weak, positive stance, accounts for about three-quarters of the instances, *opinion-offering, convincing/reasoning, reluctance to accept an idea, backchannels*, and *no-stance* represent other common stance-related uses. Combinations of acoustic-prosodic characteristics (duration, intensity, pitch) are assessed in order 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 distinguish these fine-grained stance types, and within the larger *agreement* category, positive polarity is signaled by higher pitch, lower intensity, and longer vowel duration, while greater stance strength shows higher pitch and intensity. Finally, the small set of negative 'yeahs' in the current sample is examined for comparison with the patterns described for the sample reported in Chapter 4.

5.1 Motivation

As mentioned in Chapter 4, 'yeah' can be associated with a variety of context-dependent meanings. As discourse markers, 'cue words' like 'yeah,' 'okay,' 'alright,' etc. may convey information about discourse structure and/or make a semantic contribution [40, 41, 49]. In previous studies on such cue words, prosodic variation, such as pitch accent type, has been shown to reliably distinguish discourse contributions such as backchannels from semantic

[1] Portions of this work will appear in the *(Proceedings of Interspeech 2015).*

contributions of affirmative cue words such as ‘okay’ and ‘alright’ [39, 40, 49]. 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 [40].

Given previous findings on the utility of acoustic-prosodic features in differentiating both stance strength and cue word roles, this study proposes that one or more such features differentiate stance-related uses of the word ‘yeah.’ More specifically, it is predicted that vowel duration, intensity, or pitch patterns are associated with fine-grained differences between stance-act types such as *agreement, opinion-offering, convincing/reasoning, reluctance to accept an idea,* and *backchannels.* This prediction is tested using a large sample of stance-annotated conversations taken from the ATAROS corpus, described briefly in Section 5.2, and acoustic analyses presented in Section 5.3. Findings are summarized in Section 5.4.

5.2 Corpus sample

The sample in this study is drawn from the first 20 dyads in the ATAROS corpus with two tasks annotated for both stance strength/polarity and type, as described in Chapter 2. As detailed in Chapter 6, this sample consists of 7 female-female, 3 male-male, and 10 mixed-gender dyads engaged in 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 (cf. Section 2.2.3). In this sample of 8 total hours of conversation, more than 2650 ‘yeahs’ are uttered.

As described in Section 2.3.2, 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). However, 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 assessments made for each 'yeah' independently, following the same strength/polarity scheme. For the finer-grained level of stance type annotation, each 'yeah' in the current sample inherits the stance type label of the stance act to which it belongs, as determined via the annotation procedures elaborated in Section 2.3.3.

Of all 'yeahs' in the sample, 2475 (93%) fall into the six most common categories, *agreement, no-stance, backchannels, opinion-offering, reluctance*, and *convincing* (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 5.1, 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 [40]), but lower than that observed for unstructured telephone conversations (e.g., in SWITCHBOARD [37]). The rates are also lower than those observed for the goal-oriented ICSI Meetings [69], which include both collaborative discussions and

Table 5.1: '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

reporting-oriented meetings. About half of the ‘yeahs’ in the least populated 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).

5.3 Analysis

Addressing the prediction that stance type, strength, and polarity affect acoustic-prosodic features, and building on past work that has found prosodic features useful in distinguishing stance presence and type (e.g., Chapter 4, [25, 26, 39, 40, 49]), Section 5.3.1 describes speaker-normalized measures of duration, intensity, and pitch for all ‘yeahs’ in the sample. Sections 5.3.2 and 5.3.3 examine measurable acoustic differences associated with strength and polarity within the largest stance type category, *agreement*, finishing with a more qualitative discussion of the small number of negative ‘yeahs’ in the sample.

5.3.1 Stance type

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).

Intensity

Intensity was 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 was 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)).

Following the successful use of pitch and intensity contours in Experiment 1 (Chapter 4), the smoothing-spline ANOVA plot in Figure 5.1 shows intensity contours of each stance type across word duration surrounded by shading representing 95% confidence intervals around the means [42, 100]. 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

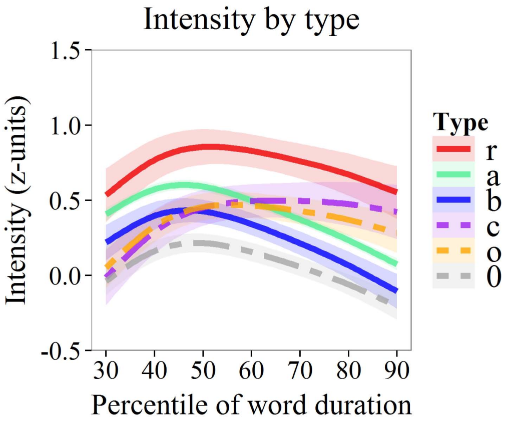


Figure 5.1: Intensity contours by stance type: all 'yeahs'

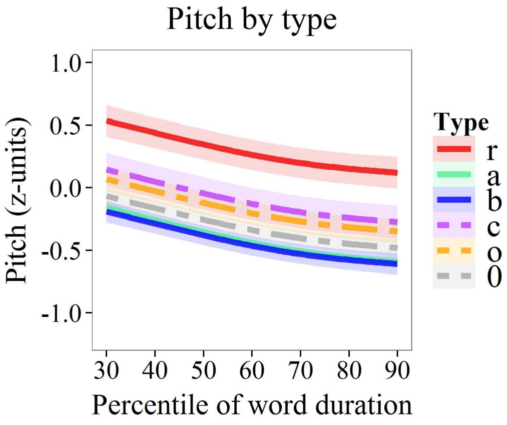


Figure 5.2: Pitch contours by stance type: all 'yeahs'

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 style: (r, a, b) in solid, (c, o, 0) in dashed. 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 *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.

Pitch

Pitch was extracted using Kaldi[2] [35] at every 10ms of word duration and then log-scaled and z-score normalized speaker-internally, similarly as 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 5.2. *Backchannels* and *agreement* have the lowest pitch, and the *backchannels* on average lack the final rise observed in other work (e.g., [40]). In addition, *reluctant* 'yeahs' have higher mean and maximum pitch than words immediately preceding them.

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

5.3.2 Stance strength

Given the prosodic differences between stance types, stance strength and polarity are examined within only the *agreement* category, the only type with sufficient tokens for further subdivision. Among the 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 5.3, in which shading

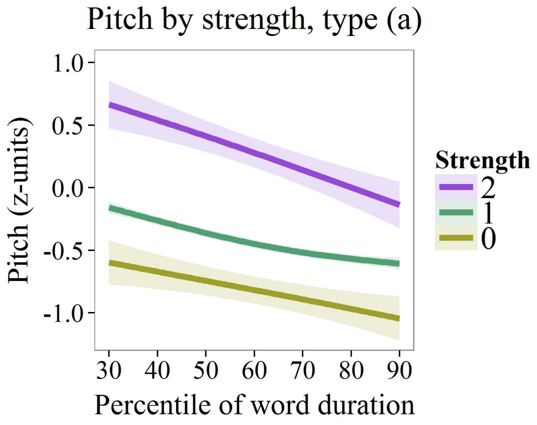


Figure 5.3: Pitch contours by stance strength: agreeing 'yeah'

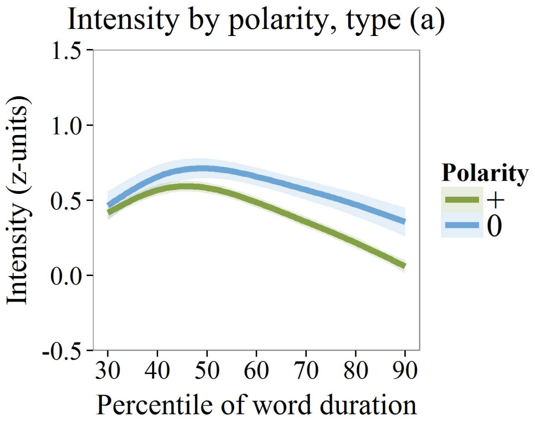


Figure 5.4: Intensity contours by stance polarity: agreeing 'yeah'

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 5.1), although weak and no-strength 'yeahs' show only slim separation.

5.3.3 Polarity

In the annotation process (cf. Section 2.3.2), 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, differences between positive and neutral 'yeahs' are investigated within the largest stance type category, *agreement*, followed by a more qualitative examination of the few negative tokens in the sample.

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 of intensity can be seen in the smoothing-spline ANOVA plot in Figure 5.4, in which mean intensity for positive *agreeing* 'yeahs' declines more sharply after word midpoint.

Negative

Here the rare but interesting negative uses of 'yeah' are discussed with more qualitative detail. As reported in Chapter 4, in a previous stage of analysis on a smaller sample of the corpus (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 (cf. Section 4.1), which were differentiated by their pitch and intensity contours: ‘yeahs’ labeled in the *tough problem* category (an expression of shared difficulty) and those under *that’s bad* (agreement with a negative assessment) group together with lower, flat pitch, while ‘yeahs’ contributing to *reluctance* (to accept a stance) and those labeled *yeah but* (preceding explanation against a stance) group with higher, curving pitch (dipping and domed, respectively), but *tough problem* and *reluctance* ‘yeahs’ show lower, relatively flat intensity, while *that’s bad* and *yeah but* ‘yeahs’ 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 dataset; 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 (cf. Section 2.3.3) 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 (cf. Chapter 4), 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, while *yeah but* and *tough problem* are lower on both measures (see Figures 5.5-5.6). In contrast to

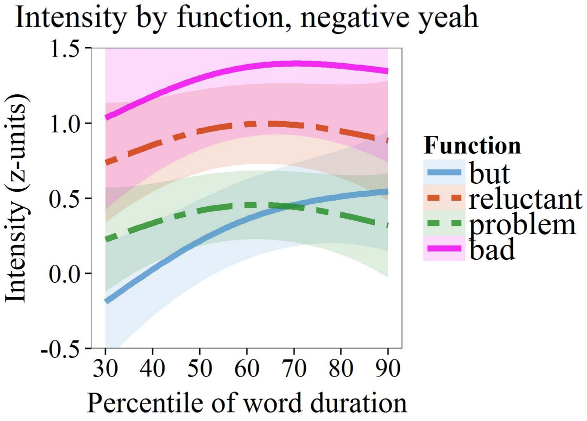


Figure 5.5: Intensity contours by stance function: negative 'yeah'

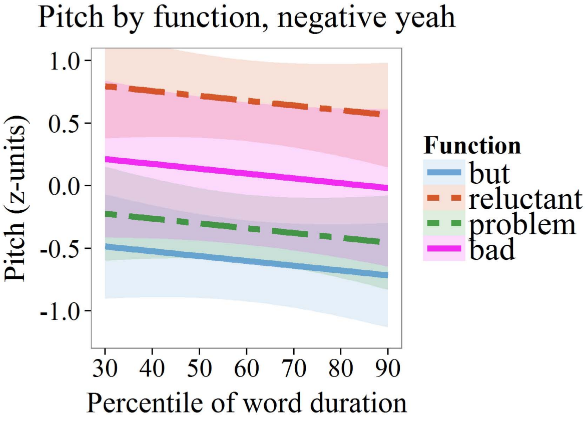


Figure 5.6: Pitch contours by stance function: negative 'yeah'

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.

5.4 Summary

In this study of stance type, function, strength, and polarity, over 2200 'yeahs' said by 40 speakers during collaborative discussions are divided into six common stance-act categories: *agreement, opinion-offering, convincing/reasoning, reluctance to accept an idea, backchannels,* and *no-stance.* The categories can be distinguished on average through a combination of prosodic cues, primarily vowel duration and intensity contours, while pitch is also useful in distinguishing stance strength and polarity. Longer vowel duration separates *agreement, reluctance,* and *backchannels* from the other categories. Intensity subdivides these clusters, with *reluctance* showing the highest intensity of all types, *agreement* moderately high intensity, and *backchannels* moderately low. *Convincing* and *opinion-offering* 'yeah' pattern together with rising intensity, while *no-stance* 'yeahs' remain low. Within the *agreement* category, moderate strength is separated from weak/no-stance by higher pitch and intensity, and positive polarity is signaled by higher pitch, lower intensity, and longer vowel duration when compared to neutral ‘yeah’. Among the few negative ‘yeahs,’ pitch and intensity may help divide four stance-related discourse functions into two groups, in contrast to the cross-cutting patterns reported in Chapter 4, but with so few tokens in each sample, this rare category is only described qualitatively.

In Chapter 6, the methods piloted in this and the previous study of ‘yeah’ (Chapter 4) are applied to an expanded sample of the corpus with access to all lexical items.

Chapter 6

EXPERIMENT 3: PHONETIC FEATURES OF STANCE IN COLLABORATIVE CONVERSATION

As the most expansive study reported here, this experiment builds on the methods and findings of Experiments 1 and 2 (Chapters 4-5) to look for signals of stance strength, polarity, and type throughout the ATAROS corpus.

6.1 Data set

The corpus sample in this study is drawn from the same 20 dyads used in Experiment 2 (the first 20 to have both stance strength/polarity and type annotation; Chapter 5), again completing the Inventory and Budget tasks, in which they arrange household items on a map of an imaginary superstore and cut expenses from an imaginary county budget, respectively (see Section 2.2.3 for more details on the tasks). Tables 6.1-6.2 show the distributions of dyads and speakers in the sample by age and sex, in the same manner as Tables 2.9-2.10 show for the entire corpus.

Table 6.1: Dyads by age and sex: 20-dyad sample

		Dyads by sex			
Group	Ages	FF	MM	MF	*Sums*
Younger	(18-32)	3	1	6	10
Middle	(38-49)	1	1	3	5
Older	(60-75)	3	1	1	5
Totals		7	3	10	20

Table 6.2: Speakers by age and sex: 20-dyad sample

		Speaker sex		
Group	Ages	F	M	*Sums*
Younger	(18-32)	12	8	20
Middle	(38-49)	5	5	10
Older	(60-75)	7	3	10
Totals		24	16	40

As detailed in Section 2.3, the tasks are manually transcribed at the utterance level, with word and phone boundaries automatically time-aligned to the audio using the Penn Phonetics Lab Forced Aligner (P2FA [108]). While Experiment 2 examines only instances of the word 'yeah,' this study includes all words in the sample, with a focus of measurement on stressed vowels in content words. The sample of 8 total hours of conversation contains over 71,300 words with a total of over 92,500 vowels. Of these vowels, 54% are uttered during the Budget task, and 57% are said by females. Contributions by speaker vary from less than 1% to 4.5% each, but they are proportional by age group, with almost half said by the younger group and about a quarter each for the middle and older groups. Following the lexical stress assigned by the pronouncing dictionary (CMUdict [102]) in the forced-aligner (P2FA [108]), 67% of vowels in the sample have primary lexical stress, 3% secondary stress, and 30% are lexically unstressed. Following the classification of words as content or function (prepositions, pronouns, articles, auxiliaries, etc.) mentioned in Section 2.4.5, 54% of vowels in the sample are in content words. Because function words and unstressed vowels are generally reduced in pronunciation in English, much of the acoustic analysis is conducted on only stressed vowels in content words; these comprise 37% of all vowels in the sample.

Phrases with poor forced-alignments (as identified during stance type annotation (cf. Section 2.3.3) are removed before acoustic analysis in order to avoid introducing errors in the automatically-extracted measurements, which are taken at points in the audio with reference to the alignments. In the current sample, about 3.5% of vowels are excluded in this process, leaving about 89,250 vowels for analysis, with all of the above proportions holding within 1%. Within the nearly 33,200 well-aligned stressed vowels in content words (henceforth 'stressed-content vowels' or SCVs), most vowel qualities are well represented, with a range of about 1500 to 4800 tokens each, except for the diphthongs /ɔɪ/ and /aʊ/, the syllabic rhotic /ɝ/ and the pre-rhotic /ɔ/, which is distinct in Northwest English, while non-pre-rhotic /ɔ/ is merged with /a/ (cf. Section 2.4.5). These each have 280-650 tokens, but this is not a problem for vowel formant analysis, as diphthongs and the rhotic are not targeted in the tasks and are not included in vowel-space plots presented here.

Also detailed in Section 2.3, utterances are hand-annotated holistically for stance strength (none, weak, moderate, strong) and polarity (positive, negative, neutral), and stance acts are identified and labeled with categories such as *opinion-offering or soliciting, (dis)agreement, convincing*, etc. Words and their vowels inherit the stance strength, polarity, and type labels applied to the spurts and stance acts to which they belong. Overall, about 40% of vowels occur in weak-strength spurts, 40% in moderate-strength, 1.5% in strong, 15% in no-stance, with the remaining 3.5% unclear. Utterances with clear stance strength (weak, moderate, strong) are also labeled for polarity; overall, about 25% of vowels in these utterances receive positive labels and 8% negative, leaving 67% in utterances with neither positive nor negative polarity. As the focus of analysis here is stressed-content vowels, Table 6.3 shows the distribution of this subset by stance strength and polarity. Note that 'x' indicates unclear polarity; these tokens are included in stance strength analysis but removed for polarity analysis. Vowels with unclear stance strength are excluded from both types of analysis.

The 24 stance type labels and label combinations with at least 100 stressed-content vowel tokens are included in the analyses of stance type presented here (Table 6.4). This helps ensure there are enough tokens in each category for reliable comparisons between types. With over 32,000 total vowels, all types in the annotation scheme (Table 2.8) are represented except *encouragement* (e). Table 6.4 shows the total number of stance acts with each label, the mean

Table 6.3: Stance strength and polarity levels: stressed-content vowels

		Polarity					
		Positive	Neutral	Negative	Unclear	NA	
Strength		(+)		(−)	(x)		*Sums*
None	(0)	-	-	-	-	5630	5630
Weak	(1)	4622	8653	217	14	-	13506
Moderate	(2)	2040	8666	1745	38	-	12489
Strong	(3)	36	213	167	4	-	420
Totals		6698	17532	2129	56	5630	32045

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 are *opinion-offering, convincing/reasoning*, and *agreement* (labels o, c, a); together, these comprise 54% of the measured stressed-content vowels. Also frequent are vowels in stretches of speech labeled here as *no-stance* (labeled 0, 24% of SCVs); these are not considered parts of stance acts, but they are included in acoustic analyses for comparison. *Opinions with solicitation or supporting reasons* (os, co) together contribute just under 9% of all SCVs, and the remaining stance types contribute less than 2% each. As mentioned in Section 2.4.4, stance act types vary 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.

6.2 Measurements

After transcription, alignment, and annotation are complete, a Praat script automatically takes measurements over all words and vowels, including pitch, intensity, and formants (f0,

Table 6.4: Stance type labels with > 100 stressed-content vowels (SCVs)

		Acts	Words/act		SCVs/act		SCVs
Stance type		Total	Mean	St.Dev.	Mean	St.Dev.	Total
o	offering opinion/suggestion	3000	5.9	3.8	2.9	1.9	7991
0	no-stance *(often not acts)*	3427	7.9	8.8	3.2	4.1	7569
c	convincing/reasoning	1564	8.7	7.0	3.9	3.1	5720
a	agreement	3292	1.9	1.6	1.4	0.9	3663
os	offer+solicit ("How about...?")	703	5.3	3.5	2.7	1.6	1786
co	opinion with reasons	267	9.0	4.8	4.1	2.4	1064
s	soliciting opinion	393	3.7	2.5	1.6	1.0	506
ot	opinion with rapport	137	7.0	4.5	3.0	2.1	386
f	softening/hesitation	345	3.0	2.1	1.4	0.9	378
cd	disagreement with reasons	92	9.6	8.5	4.2	3.3	369
t	teamwork/rapport-building	158	4.9	3.1	2.6	1.6	363
ct	reasons supporting rapport ("That's why we're so good!")	90	8.1	4.1	3.5	1.9	319
ac	agreement with reasons	82	8.1	5.1	3.8	2.3	296
cs	soliciting with reasons ("You think so because...?")	78	7.9	4.7	3.3	2.3	253
x	unclear but seems "stancey"	188	2.7	2.1	1.7	1.3	228
fo	softened offer	89	5.8	3.1	2.5	1.5	216
r	reluctance to accept a stance	184	2.2	1.8	1.3	0.7	173
b	backchannels	193	1.1	0.3	1.0	0.2	146
do	disagreement with alternative	45	8.0	3.4	3.1	1.5	139
i	strong intonation	80	2.2	1.5	2.0	1.0	120
at	agreement with rapport	72	3.2	2.4	1.9	1.2	115
d	disagreement	95	3.8	3.0	1.9	1.2	113
cr	reluctance with reasons	28	9.5	6.6	4.4	2.8	111
ao	agree and offer a new opinion	38	5.9	3.1	2.9	1.5	109

dB, F1, F2, F3) at every decile of their duration, using Praat's autocorrelation, mean energy, and LPC functions, respectively. Settings that remain fixed for all speakers include a window length of 25 ms, pitch range of 50-300 Hz, dynamic range of 30 dB, and formant range of 0-5500 Hz. Speakers are processed in batches using either 12 or 14 formant coefficients (5 or 6 formants), based on the output of a Python script that considers all speaker vowels (F1xF2) from both tasks using each formant setting and then determines which results in fewer outliers [62]. Due to the very large size of the corpus, no manual correction is done on these automatic measurements; instead, the large data set allows for a tolerance of outliers, alignment and measurement errors, which can be trimmed or ignored during analysis.

Measurements are normalized within-speaker to allow for cross-speaker comparisons. Vowel pitch and intensity are each z-score normalized using the means and standard deviations of all a speaker's measurements taken over all words in both tasks combined. That is, each raw measurement taken over deciles of vowel duration is converted to z-units by subtracting the speaker's mean and then dividing by the speaker's standard deviation:

$$z = (x - \mu)/\sigma$$

Similarly, vowel duration is z-score normalized within speaker but also within vowel quality, to account for intrinsic vowel duration differences [76, 94]. Vowel formants are normalized in R [80] using the speaker-intrinsic, vowel-extrinsic 'Nearey 2' method [71] as implemented in the phonR package [66]. To reduce the range of outliers produced from errors in automatic measurement, the highest 2.5% and lowest 2.5% of F1 and F2 measurements at each decile are trimmed from each vowel quality for each speaker. Because the R script that does this removes at least two measurements in each iteration, speaker-vowel-deciles with less than 40 measurements are trimmed by more than 5%. Overall, this results in 5.6% of all vowel measurements being trimmed, but as F1 and F2 are trimmed separately, any vowel with either measurement trimmed cannot be plotted in F1xF2 space; the result is the removal of a total of 10.2% of all vowel midpoints before the creation of the vowel space plots presented here. Removed vowels are well distributed across vowel-stress and word-function levels, so that a proportional 9.8% of stressed-content midpoints are removed by this process.

These measurements are employed in the investigation of acoustic signals of stance type, strength, and polarity, as described in the next two sections. Following the exploration of common factors known to affect acoustic-prosodic measures presented in Section 2.4.5, most measures and plots presented here combine data from both sexes and both tasks but involve only stressed vowels in content words ('stressed-content vowels' or SCVs), due to the reduction and wider variation found in function words and unstressed vowels.

6.3 Prosodic features

Following the results of Experiments 1 and 2 (Chapters 4-5), signals of stance strength, polarity, and type are sought in prosodic features, here using vowel duration and pitch and intensity at midpoint of stressed vowels in content words. Midpoint was chosen as a time point likely to be representative of the vowel as a whole, as midpoint is likely to occur during the vowel's 'steady state,' the most stable portion, and the farthest from the effects of flanking phones and edge effects caused by small inaccuracies in forced-alignment. To begin exploring the magnitude and interactions of the effects of each prosodic feature, the z-score normalized measures are submitted to a principal components analysis (PCA) using the R package ggbiplot [96]. Figures 6.1-6.3 plot the first two principal components and the measure vectors that contribute to each, overlaid with ellipses of one standard deviation around the mean of each level of stance strength, polarity, and type, respectively. The first component, on the horizontal axis, accounts for about half the variance in the data. As pitch is roughly parallel to the axis, it is the primary contributor to this dimension. The proximity of the intensity vector's angle indicates close colinearity with pitch. The second component, on the vertical axis, accounts for another third of the variance in the data; the primary contributor to this dimension is vowel duration, as its vector is nearly vertical and roughly orthogonal to pitch and intensity. The high degrees of overlap among the ellipses drawn around each level of each stance variable indicate that the levels are not easily separable using these measures alone. For example, while the wider ellipse for vowels in strong-stance utterances (level 3,

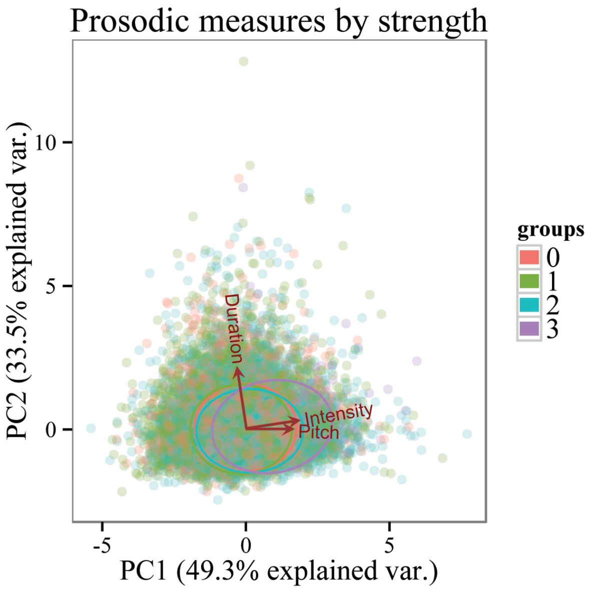


Figure 6.1: PCA plot: prosodic measures with ellipses for stance strength

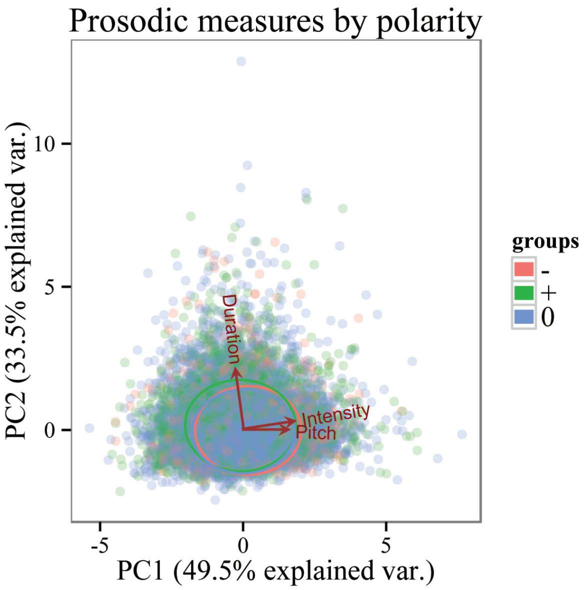


Figure 6.2: PCA plot: prosodic measures with ellipses for stance polarity

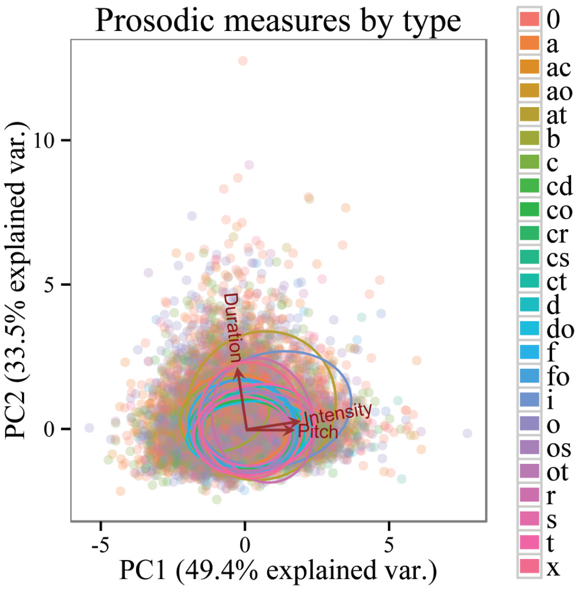


Figure 6.3: PCA plot: prosodic measures with ellipses for stance types

Figure 6.1) indicates higher means and variances in pitch and intensity, this group is still very similar to the others on these prosodic measures. The levels of polarity show even greater overlap (Figure 6.2), and only a few stance types differ noticeably (Figure 6.3); these types will be discussed below in the contexts of the measures that help distinguish them.

6.3.1 Pitch

As one of the primary contributing factors in the PCA plots described above, pitch is a useful measure for distinguishing stance strength, polarity, and type. Because exploratory examinations of pitch and intensity show that patterns hold across vowel duration, statistics are reported for these measures at vowel midpoint. One-way ANOVAs (assuming unequal variance) show significant effects for all three stance factors: strength ($F[3, 1625] = 44.5, p < 0.001$), polarity ($F[2, 4014] = 31.6, p < 0.001$), and type ($F[23, 1455] = 16.2, p < 0.001$). Welch's two-sample t-tests reveal that pitch generally increases with strength ($p < 0.001$), with the exception that no-stance vowels (label 0) are indistinguishable from moderate-strength (label 2), with higher pitch than weak-strength (label 1). All three levels of polarity differ reliably ($p < 0.001$), with negative highest and positive lowest in pitch. However, when strength and polarity labels are combined, strength is clearly the dominant factor, as is apparent when labels are arranged from low to high pitch (1+, 1-, 2-, 1, 2+, 2/0, 3-, 3, 3+). Welch's two-sample t-tests show that each combined label does not differ from its immediate neighbors, with two exceptions: moderate-negative (2-) differs from all others ($p < 0.05$), and there is a division between strong-negative (3-) and moderate/none (2/0) ($p < 0.001$). Within the stronger-stance groups (2, 3), the relative pitch heights of the polarity levels are reversed from the overall pattern, with negative utterances showing lower pitch and positive higher; the overall pattern is likely under the influence of the large number of weak-positive vowels (1+), which appear to behave as their own group, showing the lowest pitch of any type. Overall, these patterns are consistent with those found for the agreeing 'yeahs' in the sample (Experiment 2, Chapter 5), in which mean vowel pitch is higher in moderate-strength 'yeahs' than in weak and no-stance (Figure 5.3), and higher in positive 'yeahs' than neutral.

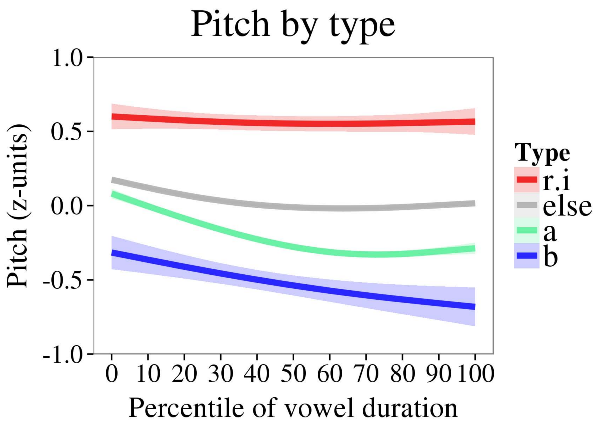


Figure 6.4: Pitch contours by stance type

Welch's two-sample t-tests show high overlap between stance types, with only a few types distinct from the others: *Reluctance to offer a stance* (r) and *strong intonation* (i) are indistinguishable with the highest pitch, *backchannels* (b) have the lowest, and *agreement* (a) dips from moderate to low ($p < 0.05$). This is a similar arrangement to the pattern found for the *agreeing* 'yeahs' in (Experiment 2, Chapter 5), in which *reluctance* (r) shows the highest mean pitch and *backchannels* and *agreement* (b, a) the lowest (Figure 5.2). All other types overlap heavily and are therefore not clearly distinguishable based on pitch at vowel midpoint. These relationships can be seen in the smoothing-spline ANOVA plot in Figure 6.4, which shows a contour connecting mean pitch for each stance type cluster identified above at each decile of vowel duration surrounded by shading representing 95% confidence intervals around the means [42, 100]. While pitch generally declines over vowel duration, *agreement* and *backchannels* (a, b) show sharper slopes. These patterns hold in words at all utterance locations, with pitch generally declining over utterance duration.

6.3.2 Intensity

Intensity at vowel midpoint is also a useful signal of stance. One-way ANOVAs (assuming unequal variance) applied to stressed-content vowels show significant effects for all three stance factors: strength ($F[3, 1977] = 283.8, p < 0.001$), polarity ($F[2, 5437] = 52.1, p < 0.001$), and type ($F[23, 1966] = 19.5, p < 0.001$). Welch's two-sample t-tests reveal that intensity generally increases with strength ($p < 0.001$), except for vowels in weak positive utterances (label 1+), which have lower intensity than no-stance and other weak-stance vowels (labels 0, 1, 1-). All three levels of polarity differ reliably ($p < 0.001$), with negative highest and positive lowest in intensity, but this appears to be a reflection of the unequal distributions between strength and polarity levels: as can be seen in Table 6.3, 69% of positive utterances have weak strength (1+), while 82% of the much smaller group of negative utterances have moderate strength (2-). When strength and polarity labels are combined, Welch's two-sample t-tests reveal the following clusters, in order of lowest to highest intensity: weak-positive (1+), weak and no-stance (0, 1, 1-), moderate neutral and negative (2, 2-), moderate positive (2+), strong (3, 3-, 3+). Thus it appears that intensity, like pitch, is more sensitive to stance strength than polarity, but weak-positive utterances behave differently, with lowered intensity. This relationship is clear in the smoothing-spline ANOVA plot in Figure 6.5, in which pitch contours over vowel duration are easily separable by stance strength but less so by polarity, and weak-positive vowels (1+) pull away with more sharply declining intensity. Overall, these patterns are consistent with those found for the *agreeing* 'yeahs' in

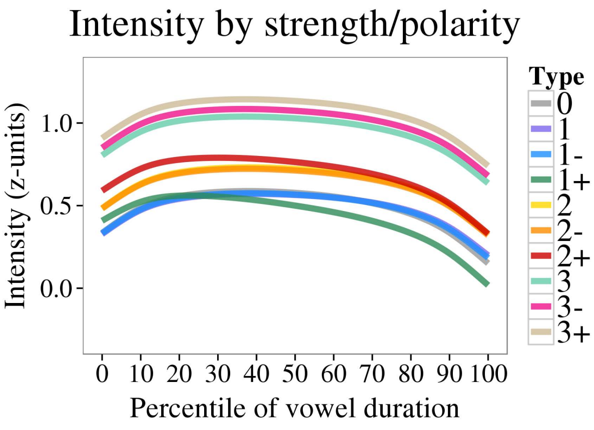


Figure 6.5: Intensity contours by stance strength/polarity

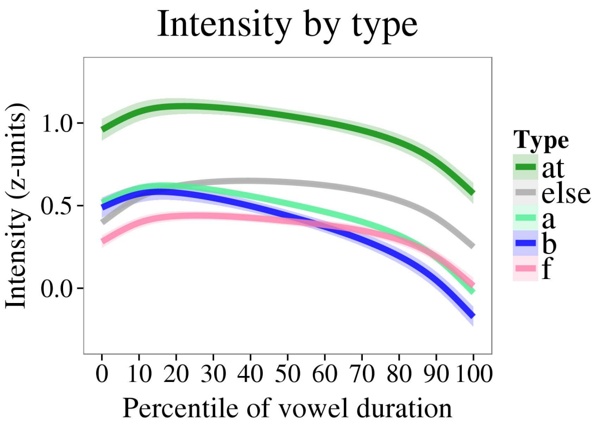


Figure 6.6: Intensity contours by stance type

the sample (Experiment 2, Chapter 5), in which mean vowel intensity is higher in moderate-strength 'yeahs' than in weak and no-stance, and lower in positive 'yeahs' than neutral (Figure 5.4).

As with pitch, Welch's two-sample t-tests show high overlap between stance types, with only a few types appearing more distinct from others. *Agreement with rapport* (at) has the highest intensity and differs significantly from all other types except *strong intonation* (i) ($p < 0.01$). Its also drops less at the ends of utterances than other types. *Stance-softening or hesitation* (f) has the lowest intensity and overlaps only with *backchannels* (b), the next highest, which in turn overlaps with the next highest, *agreement* (a) ($p < 0.05$). Both *agreement* and *backchannels* (a, b) drop more sharply over vowel duration than other types. All other types overlap heavily and are therefore not clearly distinguishable based on intensity at vowel midpoint. These patterns can be seen in the smoothing-spline ANOVA plot in Figure 6.6, which shows a contour connecting mean intensity at each decile of vowel duration for each stance type cluster identified above [42]. While intensity generally declines over vowel duration (with drops at the edges, as expected near flanking consonants or silence), *agreement* and *backchannels* (a, b) show sharper slopes, similar to their pattern for pitch. These relationships are consistent with but more robust than those found for the *agreeing* 'yeahs' in Experiment 2 (Chapter 5), in which *reluctance* (r) shows the highest mean vowel intensity and *backchannels* and *agreement* (b, a) shows lower, more sharply declining slopes (Figure 5.1). The patterns hold in words at all utterance locations, with intensity generally declining over utterance duration.

6.3.3 Vowel duration

Finally, vowel duration also plays a role. One-way ANOVAs (assuming unequal variance) applied to stressed-content vowels show significant effects for all three stance factors: strength ($F[3, 1640] = 25.5, p < 0.001$), polarity ($F[2, 4866] = 72.5, p < 0.001$), and type ($F[23, 1967] = 31.3, p < 0.001$). Welch's two-sample t-tests reveal that vowel duration decreases with each increase in strength level ($p < 0.001$), with the exception of the rare strong-stance vowels (label 3), which are more variable and do not differ from any other strength level. For stance polarity, Welch's two-sample t-tests reveal two clusters, with positive utterances displaying longer stressed vowel duration ($p < 0.001$) than neutral and negative, which do not differ reliably. This is the same pattern found for the *agreeing* 'yeahs' in the sample, as reported in Experiment 2 (Chapter 5). With strength and polarity labels combined, polarity appears to be the primary divisor, with positive weak/moderate and no-stance vowels (labels 1+, 2+, 0) clustering with longer vowel duration than most neutral and negative vowels (labels 1, 2, 2-, 3-). Weak negative vowels (1-) are similarly short but only differ from weak positive (1+). Strong neutral vowels (3) cluster with the longer positive group, and while strong positive vowels (3+) are also long, they are more variable as a group and do not differ from any other. Thus, while the combination of strength and polarity has a complicated effect on vowel durations, which reflect speaking rate, it appears that positive stances are said more slowly and stronger stances more quickly, with the strongest stances being too variable and perhaps too rare to show a clear pattern.

For stance type, Welch's two-sample t-tests again show high overlap between types. However, a few types appear more distinct from others: *backchannels, agreement with rapport*, and *strong intonation* (at, b, i) have some of the longest vowel durations and are only indistinguishable from each other and *unclear stance* (x), which also overlaps *agreement* (a) and five other types. *Agreement* (a) also has longer vowel durations and is only indistinguishable from *unclear* (x) and two other types (fo, r). *No-stance* (0) utterances have only slightly longer-than-average vowels but overlap with only three other types (fo, r, cs). All other types overlap heavily and are therefore not clearly distinguishable based on vowel duration.

6.3.4 Combined prosodic patterns

Following the patterns of each measure above, a few of the stance types can be differentiated with a combination of prosodic features. *Agreement* (a), one of the most frequent categories, shows longer vowel duration and moderately low pitch and intensity which both dip over the course of stressed-content vowels. *Backchannels* (b), one of the least frequent types in the corpus, also show long vowel duration and low-dropping intensity, but their pitches remain low throughout vowel duration. *Reluctance to accept a stance* (r) and *strong intonation* (i), also infrequent, show high pitch, the latter also with long vowel duration. *Agreement with rapport* (at) stands out with the highest intensity and longest vowel duration, and *stance-softening/hesitation* (f) shows the lowest intensity.

The same prosodic measures also combine to help differentiate levels of stance strength and polarity. Successive levels of strength are best distinguished by increases in both pitch and intensity, while positive polarity is signaled by longer vowel duration. In combining all three measures, weak-positive utterances (1+) stand out as having the longest vowels with the lowest pitch and intensity; this group shows the same patterns as the *agreement* category mentioned above (a), as the majority of (66%) *agreeing* stance acts (a) occur in weak-positive utterances (1+), and nearly half (47%) of vowels in weak-positive utterances (1+) contribute to *agreement* (a), with another 5% involved in a combination of types which include *agreement* (ac, ae, aet, af, afo, ai, ao, ar, as, at).

6.4 Vowel space

While prosodic measures are helpful in distinguishing stance features, vowel spaces appear largely unaffected. Using mean formant measures (F1, F2) taken at midpoint of stressed vowels in content words ('stressed-content vowels' or SCVs), there is no clear pattern for stance strength, polarity, or type. Figures 6.7-6.8 show mean vowel positions by strength and polarity; while polarity clearly has no discernible effect, strong-stance vowels (label 3 in Figure 6.7) appear to be shifted downward. This effect seems driven by males, as seen in Figure 6.9 compared to females in Figure 6.10, who show much less shifting. However, as there are only 389 strong-stance vowels (1.3% of stressed-content vowels with discernible stance strength, two-thirds said by females), which display wide variation suggestive of a high rate of automatic-measurement error, it is difficult to say whether this shifting of vowels in all locations in the vowel space (front/back, low/high) is a reliable indicator of strong stance.

Plots including all 24 stance-act labels listed in Table 6.4 are very cluttered in appearance

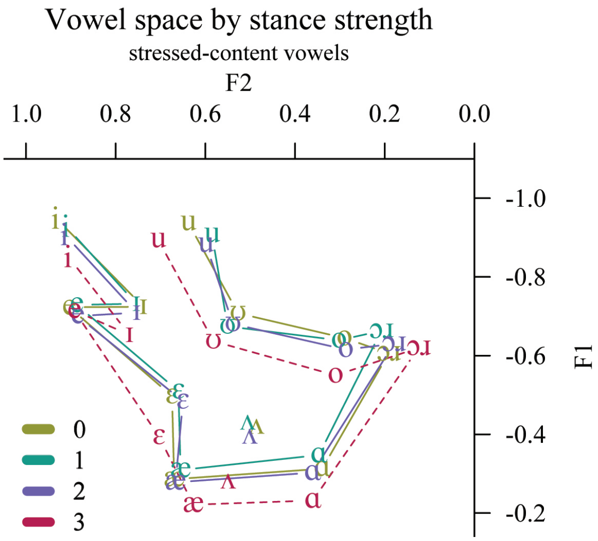


Figure 6.7: Vowel space by strength: stressed-content vowels (Nearey2-normalized)

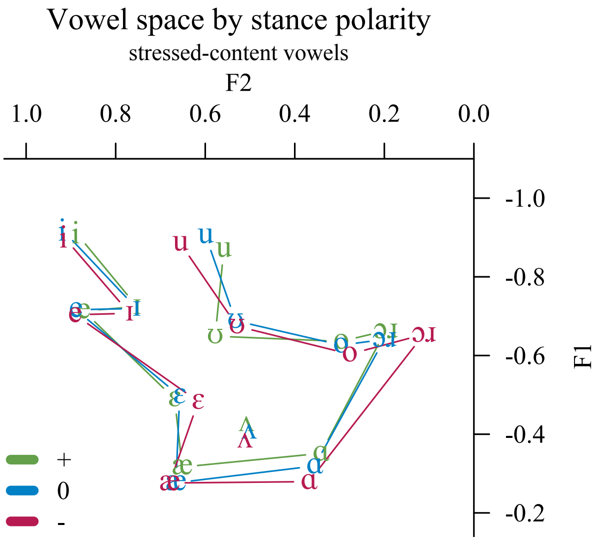


Figure 6.8: Vowel space by polarity: stressed-content vowels (Nearey2-normalized)

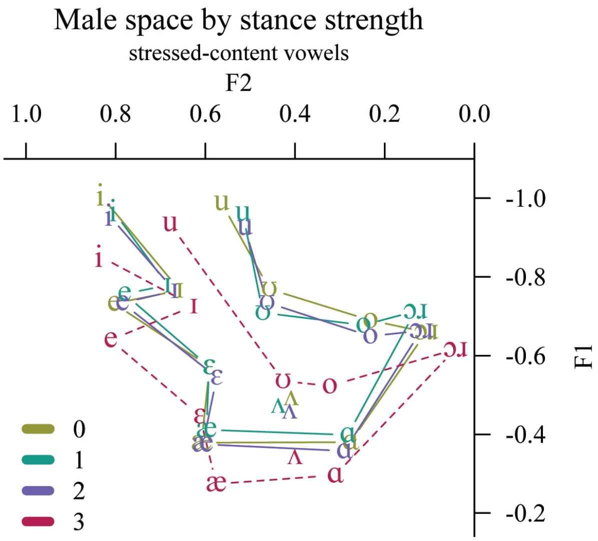


Figure 6.9: Vowel space by strength: males' stressed-content vowels (Nearey2-normalized)

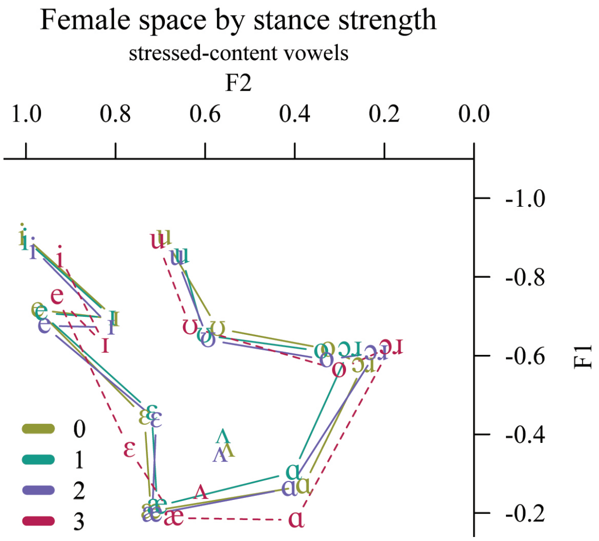


Figure 6.10: Vowel space by strength: females' stressed-content vowels (Nearey2-normalized)

and difficult to interpret, so a plot was made for each stance type showing all label combinations which include that type (e.g., d, cd, do on a plot of *disagreeing* vowels). However, visual inspections of these plots revealed no clear patterns for any stance type. Plots were also made showing single stance-type labels (a, b, etc.) and various groupings of labels, again with no clear patterns. Thus, with the current measures, very little can be said about the effect of stance type on vowel space.

6.5 Discussion

In this study of a large sample of over 32,000 stressed vowels in content words said by 40 speakers, prosodic measures are shown to be useful signals of stance strength, polarity, and type. Pitch and intensity account for half the variance in the data and are most associated with differences in stance strength and type: both increase with stance strength, and they help distinguish several stance-act types. *Reluctance* and *expressive intonation* (r, i) have very high pitch, *backchannels* (b) very low, and *agreement* (a) low-dipping; the latter two also show sharply-dropping intensity, with *backchannels* lower overall. *Stance-softening/hesitation* (f) shows the lowest intensity and *rapport-building agreement* (at) the highest. While most of these types also have longer vowels, vowel duration does not help differentiate within them; however, longer vowel duration is the key distinguisher of positive polarity. Finally, *weak-positive agreement* (a,1+) stands out with the longest vowels and lowest pitch and intensity. While vowel formants are examined, no meaningful associations are found between mean vowel space arrangements and stance type, strength, or polarity. Table 6.5 summarizes these results.

These findings support the prediction that information about stance is carried in the acoustic speech signal. Intuitively, it makes sense that variations in prosodic features 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 its meaning from statement to question, scolding to incredulous, genuine to sarcastic, and so on, but in naturally-occurring speech, such well-

Table 6.5: Summary of results

Stance feature/type		Pitch	Intensity	Duration
	Strength	increases with strength levels	increases with strength levels	–
	Polarity	–	–	positive longer
r; i	reluctance; intonation	very high	–	long
at	agreement+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	–

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 begins to identify a few small components as they are 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 hints of the broader pattern to emerge. Again, it makes sense intuitively that stronger stances have higher pitch and intensity, indicators of increased effort and investment encoded in their delivery; that backchannels and weak agreement are quiet and low-pitched; that rapport-building agreement is delivered like a long, loud cheer; that downplaying a stance is done quietly; and it is easy to imagine the long, high-pitched contour that expresses reluctance to accept an idea without outright rejection. Such intuitively valid findings form a solid foundation for expansion into both broader and more detailed acoustic investigations. For example, a phrasal-level examination should take into account not only the prosodic contour of the entire phrase, but also the more specific meanings associated with local deviations from the expected contour. This line of inquiry involves investigation of known elements that influence pronunciation (e.g., information structure, lexical predictability, discourse factors) so that unexpected deviations can be attributed to as-yet understudied factors such as stance. This was the approach taken in the precursor to the current project [25, 26], and with the availability of the large ATAROS corpus designed to be useful for just such techniques, it will continue to be a focus in future work.

Chapter 7

CONCLUSION

To conclude the three studies presented here, this chapter summarizes their results, discusses the contributions of the experiments and the ATAROS corpus, and suggests avenues for future work.

7.1 Summary of results

Support is found in all three experiments for the prediction that stance is signaled acoustically, particularly using prosodic measures.

In Experiment 1 (Chapter 4), content analysis identifies four discourse functions within a small sample of instances of the word 'yeah' that contribute to negative stances; these functions are then seen to differ by a combination of the shapes of their pitch and intensity contours.

In Experiment 2 (Chapter 5), the six most common stance types found in utterances containing over 2200 'yeahs' are divisible with a combination of vowel duration and intensity measures taken on the 'yeahs.' Specifically, *reluctant, agreeing*, and *backchanneling* 'yeahs' show longer vowel duration than *convincing, opinion-offering*, and *no-stance*, and successively decreasing mean vowel intensity separates the members of each group. Within *agreeing* 'yeahs,' higher mean pitch and intensity separate moderate-strength stance from weak/no-stance, and positive stance is signaled by longer vowels, higher pitch, and lower intensity.

In Experiment 3 (Chapter 6), similar patterns are found within the larger sample of over 32,000 stressed vowels of content words said by 40 speakers engaged in two collaborative tasks (the same speakers/tasks used in Experiment 2). However, here pitch and intensity are more informative than vowel duration. Both pitch and intensity increase with stance strength, while positive polarity is distinguished primarily by longer vowel duration. As for notable stance types, *weak-positive agreement* appears to behave as a separate group, with the longest vowels and lowest pitch and intensity. In general, *agreement* displays long vowel durations, low-dipping pitch, and dropping intensity; *backchannels* show similar patterns but are lower in both pitch and intensity. *Reluctance* shows very high pitch, along with *strongly-expressive intonation*, which also involves long vowel duration. Finally, *stance-softening* has low intensity, and *rapport-building agreement* has very high intensity. Formants are examined at vowel midpoint, but no clear signals of stance are found in vowel space arrangements.

7.2 Contributions

Taken together, the three studies presented here provide an initial sketch of the prosodic cues to stance, the ways in which components like pitch, 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, for use in both automatic detection and human-interactive production. Given that many other types of information – social, indexical, 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. For example, stance-taking behavior may vary within different social groups, when talking to different audiences, or in different social situations. Manipulation of stance is likely a component of identity-expression, group affiliation, politeness, and power dynamics, as well as an important feature of social activities such as collaboration, negotiation, and persuasion. Cross-cutting social factors, stance is also likely to interact with information structure and discourse factors, such as the pronunciation of new vs. given information, as found in [25, 26], focus contrast, turn-taking cues, etc.

While engineers and computer scientists commonly take ‘big-data’ approaches to speech-signal processing, theoretical and experimental linguists have traditionally relied on much smaller bodies of hand-measured or human-supervised data. Although the trade-off between accuracy and statistical power must always be considered in study design, the work presented here provides an example of how a complex linguistic question that has been difficult to investigate on a small scale can benefit from the computational examination of a large amount of data extracted relatively quickly and with little human intervention, allowing for a wider and potentially more generalizable view of the behavior from a variety of angles. The studies presented here are primarily socio-phonetic in nature, combining qualitative methods of content analysis with quantitative acoustic measurements, while others produced by the ATAORS team fit more in the realm of computer science, with computational linguists and speech-signal-processing engineers working toward improvements in automatic sentiment detection. In working together, each discipline has benefited from the other's expertise, and the partnership can serve as a model for future cooperation between experts in fields who study similar questions from different perspectives.

Another major contribution of this dissertation is the ATAROS corpus itself. As a targeted body of naturalistic interaction designed specifically for linguistic study, it has high audio quality, detailed annotations, and similar amounts of speech in each setting from each of 68 Northwestern English speakers with known demographics. It offers a large sample of Pacific Northwest English, a relatively young and understudied dialect, in more free-flowing conversation than what is obtained by many linguistic interview procedures. Sets of target items with varying phonetic compositions are shared across tasks, enabling direct comparisons across conversational settings designed to encourage differing levels of involvement and types of stance-taking behaviors. The collaborative task designs and stance annotation schema provide replicable and expandable models for employment in similar endeavors. The annotation schema in particular represent the collection of many descriptions of stance in conversation/discourse analytic approaches in an attempt to standardize their application to wider domains. While such annotation is subjective by nature, variation in decision points can be mitigated with detailed yet flexible categorization to yield high inter-rater agreement. Importantly, nearly all aspects of the corpus – including audio recordings, demographic information, time-aligned transcriptions and annotations, labeling schema, elicitation and analysis materials – are available to researchers outside the host institution, through the Linguistic Phonetics Lab at the University of Washington (`http://depts.washington.edu/phonlab/projects.htm`). Others at UW have already begun using the corpus for work on such wide-ranging topics as sarcasm, phonation, and disfluency, and in the future, the team hopes to expand into both audio and video recording of pairs of friends, larger groups, and/or speakers from other dialects.

7.3 Future work

As some of the first work to report acoustic signals of stance-taking, the measures and methods of labeling stance used here serve as a springboard for expansion into a variety of stance-related investigations. Future work will expand beyond unitary vowel measurements to include more complex prosodic descriptions of the corpus, for example by examining contours and timing over whole phrases or utterances. Intuitively, some of the stance types classified here may have identifiable 'tunes,' or prosodic pitch accent contours akin to those used in the ToBI system [77] to distinguish speech acts like declaratives, yes-no questions, etc. For example, a prototypical example of the polite alternative to disagreement *reluctance to accept* might be a very elongated "Well..." with high level pitch and a pinched voice quality, followed by a long pause and preceded by a pause and/or hesitation noises.

A wider view of higher-level conversational forces will also consider effects of entropy and predictability on pronunciation. At a lexical or syntactic level, this can include measures of lexical frequency and models of perplexity to determine how unusual or expected a word is in context, and therefore how carefully it is likely to be articulated [2, 5, 53]. Such predictions will then be compared to acoustic measures indicative of degrees of articulatory effort or precision (including those used here). In short, if material that the listener expects to be reduced in pronunciation (based on syntactic, information-structure, or discourse factors) is instead hyperarticulated, some other social or attitudinal meaning must be intended [26].

Similarly, future work will investigate social factors such as the ages and genders of speakers and partners, as well as rapport or power dynamics within dyads. For example, what acoustic differences are there between dyads who seem to be getting along compared to those who remain distant? How do speakers who make more suggestions or whose decisions are accepted more often differ from those who tend to agree or acquiesce? Audience design approaches (e.g., [4, 34]) predict style differences in response to attributes of speakers' partners, the researchers, and imagined third-party observers who may listen to the recordings later. If given the opportunity to expand recording, such speaker/partner traits as age differentials, familiarity, and English fluency can be manipulated, or power dynamics could be imposed using confederates. The study design used here provides an example of how naturalistic conversation can be obtained in a laboratory setting; tasks could be added to increase the range of styles elicited, whether more formal (e.g., via reading tasks), informal (e.g., by leaving friends to converse after the researcher makes an excuse to leave the lab), or personally investing (e.g., by encouraging opinionated discussion of hot-button issues).

Finally, no investigation of a human language behavior is complete without considering both production and perception. Judgment-perception experiments are currently being designed in which the features found to be indicative of stance in production will be manipulated both independently and in combination to discover the degrees to which each affects the perception of stance-expression by human listeners. For example, pitch, intensity, and vowel duration might be varied in steps using edited or synthetic speech; listeners might hear two variants of the same lexical material and then indicate which they perceive as expressing the stronger stance. To explore relationships to social factors, acoustic manipulations could be paired with differing faces or other information on the 'speaker' in a matched-guise experiment that could shed light on listeners' expectations and the social meanings which can be conveyed using the same acoustic cues.

7.4 Final remarks

In sum, this project provides a large, high-quality audio corpus of Pacific Northwest English and one of the most extensive phonetic investigations of stance-taking to date. In examining over 30,000 automatically-measured vowels together, it combines a wide-scope approach with very local measurements. This results in broad observations at a fairly coarse level of linguistic analysis, namely that features of stance strength, polarity, and type can be differentiated with combinations of prosodic features. This leaves ample room for explorations of patterns in more detailed linguistic divisions and inspection of acoustic differences at more local levels.

BIBLIOGRAPHY

[1] John L. Austin. *How to do things with words*, volume 367. Oxford University Press, 1975.

[2] Matthew Aylett and Alice Turk. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1):31–56, 2004.

[3] Kara Becker, Anna Aden, Katelyn Best, Rena Dimes, Juan Flores, and Haley Jacobson. Keep Portland weird: Vowels in Oregon English, 2013. Paper presented at New Ways of Analyzing Variation (NWAV 42), Pittsburgh.

[4] Alan Bell. Language style as audience design. *Language in Society*, 13(2):145–204, 1984.

[5] Alan Bell, Jason M. Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1):92–111, 2009.

[6] Štefan Beňuš, Agustín Gravano, and Julia Hirschberg. The prosody of backchannels in American English. In *Proceedings of the 16th International Congress on Phonetic Sciences (ICPhS)*, 2007.

[7] Douglas Biber and Edward Finegan. 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, 1989.

[8] Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. *Longman grammar of spoken and written English.* Longman, 1999.

[9] Charles Boberg. Regional phonetic differentiation in Standard Canadian English. *Journal of English Linguistics*, 36(2):129–154, 2008.

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

[11] Penelope Brown and Stephen Levinson. Universals in language usage: Politeness phenomena. In *Questions and Politeness: Strategies in Social Interaction*, pages 56–311. 1978.

[12] U.S. Census Bureau. State and county quickfacts: King County, Washington, 2010. Retrieved May 3, 2015 from `http://quickfacts.census.gov/qfd/states/53/53033.html`.

[13] Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska, Iain McCowan, Wilfried Post, Dennis Reidsma, and Pierre Wellner. The AMI meeting corpus. In *Proceedings of the Measuring Behavior Symposium on Annotating and Measuring Meeting Behavior*, 2005.

[14] Jean Carletta, Stephen Isard, Gwyneth Doherty-Sneddon, Amy Isard, Jacqueline C. Kowtko, and Anne H. Anderson. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–31, 1997.

[15] Herbert H. Clark and Susan E. Haviland. Comprehension and the given-new contract. In Roy Freedie, editor, *Discourse Production and Comprehension*, Discourse Processes: Advances in Research and Theory, pages 1–40. Lawrence Erlbaum Associates, Hillsdale, NJ, 1977.

[16] Sandra Clarke, Ford Elms, and Amani Youssef. The third dialect of English: Some Canadian evidence. *Language Variation and Change*, 7(02):209–228, 1995.

[17] Cynthia G. Clopper, David B. Pisoni, and Kenneth de Jong. Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America (JASA)*, 118(3):1661–1676, 2005.

[18] Jeff Conn. It's not all rain and coffee: An investigation into the western dialect of Portland, Oregon, 2002. Paper presented at New Ways of Analyzing Variation (NWAV 31), Stanford, California.

[19] Susan Conrad and Douglas Biber. Adverbial marking of stance in speech and writing. In Susan Conrad and Douglas Biber, editors, *Evaluation in text: Authorial stance and the construction of discourse*, pages 56–73. Oxford University Press, 2000.

[20] John W. Du Bois. The stance triangle. In *Stancetaking in discourse: Subjectivity, evaluation, interaction*, pages 139–184. John Benjamins, Amsterdam, 2007.

[21] Penny Eckert. California vowels. `http://www.stanford.edu/~eckert/vowels.html`, 2005.

[22] Robert Englebretson. Stancetaking in discourse: An introduction. In *Stancetaking in discourse: Subjectivity, evaluation, interaction*, pages 1–26. John Benjamins, Amsterdam, 2007.

[23] Norman Fairclough. *Analysing discourse: Textual analysis for social research.* Psychology Press, 2003.

[24] David W. Foster and Robert J. Hoffman. Some observations on the vowels of Pacific Northwest English (Seattle area). *American Speech*, 41(2):119–122, 1966.

[25] Valerie Freeman. Using acoustic measures of hyperarticulation to quantify novelty and evaluation in a corpus of political talk shows. Master's thesis, University of Washington, 2010.

[26] Valerie Freeman. Hyperarticulation as a signal of stance. *Journal of Phonetics*, 45:1–11, 2014.

[27] Valerie Freeman. The prevelar vowel system in Seattle. Poster presented at the American Dialect Society (ADS) Annual Meeting, Portland, OR, 2015.

[28] Valerie Freeman, Julian Chan, Gina-Anne Levow, Richard Wright, Mari Ostendorf, and Victoria Zayats. Manipulating stance and involvement using collaborative tasks: An exploratory comparison. In *Proceedings of the 15th Annual Conference of the International Speech Communication Association (Interspeech 2014)*, 2014.

[29] Valerie Freeman, Julian Chan, Gina-Anne Levow, Richard Wright, Mari Ostendorf, Victoria Zayats, Yi Luan, Heather Morrison, Lauren Fox, Maria Antoniak, and Phoebe Parsons. ATAROS technical report 1: Corpus collection and initial task validation. Technical report, University of Washington Linguistic Phonetics Lab, 2014. Available online: `http://depts.washington.edu/phonlab/projects.htm`.

[30] Valerie Freeman, Gina-Anne Levow, and Richard Wright. Phonetic marking of stance in a collaborative-task spontaneous-speech corpus. Poster presented at the 167th Meeting of the Acoustical Society of America (ASA), Providence, RI, May 5-9, 2014.

[31] Valerie Freeman, Gina-Anne Levow, Richard Wright, and Mari Ostendorf. Manipulating stance and involvement using collaborative tasks: An exploratory comparison. In *Proceedings of the 16th Annual Conference of the International Speech Communication Association (Interspeech 2015)*, 2015.

[32] Valerie Freeman, Richard Wright, and Gina-Anne Levow. ARIES: A corpus for the acoustic analysis of stance. Presentation at the Northwest Linguistics Conference (NWLC), University of Washington, Seattle, Apr. 7-8, 2012.

[33] Valerie Freeman, Richard Wright, and Gina-Anne Levow. The prosody of negative 'yeah'. In *The LSA Annual Meeting Extended Abstracts (ExtAbs)*, 2015.

[34] Susan R. Fussell and Robert M. Krauss. Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, 19(6):509–525, 1989.

[35] Pegah Ghahremani, Bagher BabaAli, Daniel Povey, Korbinian Riedhammer, Jan Trmal, and Sanjeev Khudanpur. A pitch extraction algorithm tuned for automatic speech recognition. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.

[36] AKG Acoustics GmbH. *MicroMic: AKG C 520/C 520 L*, n.d.

[37] John Godfrey, Edward Holliman, and Jane McDaniel. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 517–520, 1992.

[38] Matthew Gordon. The West and Midwest: Phonology. In Edgar W. Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie, and Clive Upton, editors, *Handbook of varieties of English*. Mouton de Gruyter, New York, 2004.

[39] Agustín Gravano, Štefan Beňuš, Héctor Chávez, Julia Hirschberg, and Lauren Wilcox. On the role of context and prosody in the interpretation of 'okay'. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL)*, pages 800–807, 2007.

[40] Agustín Gravano, Julia Hirschberg, and Štefan Beňuš. Affirmative cue words in task-oriented dialogue. *Computational Linguistics*, 38(1):1–39, 2012.

[41] Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

[42] Chong Gu. *Smoothing Spline ANOVA Models*. Springer, New York, NY, 2002.

[43] Pentti Haddington. Stance taking in news interviews. *SKY Journal of Linguistics*, 17:101–142, 2004.

[44] Robert Hagiwara. Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America*, 102(1):655–658, 1997.

[45] Michael A. K. Halliday. *Functional Grammar.* Edward Arnold, London, 1994.

[46] Rom Harré and Luk Van Langenhove. Varieties of positioning. *Journal for the Theory of Social Behaviour*, 21(4):393–407, 1991.

[47] John Heritage and Geoffrey Raymond. The terms of agreement: Indexing epistemic authority and subordination in talk-in-interaction. *Social Psychology Quarterly*, 68(1):15–38, 2005.

[48] Dustin Hillard, Mari Ostendorf, and Elizabeth Shriberg. Detection of agreement vs. disagreement in meetings: Training with unlabeled data. In *Proceedings of HLT-NAACL Conference*, Edmonton, Canada, 2003.

[49] Julia Hirschberg and Diane Litman. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, 19(3):501–530, 1993.

[50] Susan Hunston and Geoff Thompson. Evaluation: An introduction. In Susan Hunston and Geoff Thompson, editors, *Evaluation in text: Authorial stance and the construction of discourse*, pages 1–27. Oxford University Press, New York, 2000.

[51] Jennifer K. Ingle, Richard Wright, and Alicia Wassink. Pacific Northwest vowels: A Seattle neighborhood dialect study. *Journal of the Acoustical Society of America (JASA)*, 117(4):2459–2459, 2005.

[52] Alexandra Jaffe. Introduction: The sociolinguistics of stance. In Alexandra Jaffe, editor, *Stance: Sociolinguistic Perspectives*, Oxford Studies in Sociolinguistics. Oxford University Press, New York, 2009.

[53] Daniel Jurafsky, Alan Bell, Michelle Gregory, and William D. Raymond. Effect of language model probability on pronunciation reduction. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Salt Lake City, Utah, 2001.

[54] Elise Kärkkäinen. Stance taking in conversation: From subjectivity to intersubjectivity. *Text & Talk - An Interdisciplinary Journal of Language, Discourse Communication Studies*, 26(6):699–731, 2006.

[55] Robert Kennedy and James Grama. Chain shifting and centralization in California vowels: An acoustic analysis. *American Speech*, 87(1):39–56, 2012.

[56] Paul Kingsbury, Stephanie Strassel, Cynthia McLemore, and Robert McIntyre. CALLHOME American English Transcripts LDC97T14. `https://catalog.ldc.upenn.edu/LDC97T14`, 1997.

[57] Paul Kockelman. Stance and subjectivity. *Journal of Linguistic Anthropology*, 14(2):127–150, 2004.

[58] William Labov. *Language in the inner city.* University of Pennsylvania, Philadelphia, 1972.

[59] William Labov. *Principles of Linguistic Change, Internal Factors*, volume 1. Wiley-Blackwell, Malden, MA, 1994.

[60] William Labov, Sharon Ash, and Charles Boberg. *The Atlas of North American English: Phonetics, phonology and sound change.* Walter de Gruyter, Berlin, 2006.

[61] Gina-Anne Levow, Valerie Freeman, Alena Hrynkevich, Mari Ostendorf, Richard Wright, Julian Chan, and Trang Tran. Recognition of stance strength and polarity in spontaneous speech. In *Proceedeings of the 5th IEEE Workshop on Spoken Language Technology (SLT)*, 2014.

[62] Yi Luan, Richard Wright, Mari Ostendorf, and Gina-Anne Levow. Relating automatic vowel space estimates to talker intelligibility. In *Proceedings of the 15th Annual Conference of the International Speech Communication Association (Interspeech 2014)*, 2014.

[63] Herbert W. Luthin. The story of California (ow): The coming-of-age of English in California. In Keith M. Denning, Sharon Inkelas, Frances C. McNair-Knox, and John R. Rickford, editors, *Variation in Language, NWAV-XV at Stanford: Proceedings of the 15th Annual Conference on New Ways of Analyzing Variation*, pages 312–24. Department of Linguistics, Stanford University, Stanford, CA, 1987.

[64] M-Audio. *ProFire 610*, 2008.

[65] James R. Martin and Peter R. White. *The language of evaluation: Appraisal in English.* Palgrave Macmillan, New York, 2005.

[66] Daniel R. McCloy. phonR: Tools for phoneticians and phonologists. [R package version 1.0-3], 2015. Available online: `https://github.com/drammock/phonR`.

[67] Birch Moonwomon. Truly awesome: ('open o') in California English. In Keith M. Denning, Sharon Inkelas, Frances C. McNair-Knox, and John R. Rickford, editors, *Variation in Language, NWAV-XV at Stanford: Proceedings of the 15th Annual Conference on New Ways of Analyzing Variation*, pages 325–336. Department of Linguistics, Stanford University, Stanford, CA, 1987.

[68] Franc Morales. Academic resources: Function words. Sequence Publishing, 2015. Retrieved Apr. 15, 2015 from `http://www.sequencepublishing.com/academic.html`.

[69] Nelson Morgan, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Adam Janin, Thilo Pfau, Elizabeth Shriberg, and Andreas Stolcke. The meeting project at ICSI. In *Proceedings of Human Language Technologies*, 2001.

[70] Gabriel Murray and Giuseppe Carenini. Detecting subjectivity in multiparty speech. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, pages 2007–2010, 2009.

[71] Terrance M. Nearey. *Phonetic feature systems for vowels.* Doctoral dissertation, University of Alberta, 1978. Reprinted by the Indiana University Linguistics Club.

[72] Richard Ogden. Phonetics and social action in agreements and disagreements. *Journal of Pragmatics*, 38(10):1752–1775, 2006.

[73] John J. Ohala and Brian W. Eukel. Explaining the intrinsic pitch of vowels. In Robert Channon and Linda Shockey, editors, *In Honor of Ilse Lehiste*, pages 207–215. Foris, Dordrecht, 1987.

[74] Mari Ostendorf and Sangyun Hahn. A sequential repetition model for improved disfluency detection. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, 2013.

[75] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86, 2002.

[76] Gordon E. Peterson and Ilse Lehiste. Duration of syllable nuclei in english. *Journal of the Acoustical Society of America (JASA)*, 32(6):693–703, 1960.

[77] John F. Pitrelli, Mary E. Beckman, and Julia Hirschberg. Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 1994.

[78] Ellen F. Prince. Toward a taxonomy of given-new information. In Peter Cole, editor, *Radical pragmatics*, pages 223–255. Academic Press, New York, 1981.

[79] Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. *A Comprehensive Grammar of the English Language*. Longman, New York, 1985.

[80] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015. `http://www.R-project.org/`.

[81] Stephan Raaijmakers, Khiet Truong, and Theresa Wilson. Multimodal subjectivity analysis of multiparty conversation. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 466–474, Honolulu, 2008.

[82] Scouting Web: Online resources for scouting volunteers. Survival: A simulation game. `http://scoutingweb.com/scoutingweb/SubPages/SurvivalGame.htm`, 2014.

[83] John M. Riebold. Please merge ahead: The vowel space of Pacific Northwestern English. Paper presented at the Northwest Linguistics Conference (NWLC), University of Washington, Seattle, Apr. 7-8, 2012.

[84] John M. Riebold. Language change isn't only skin deep: Inter-ethnic contact and the spread of innovation in the Northwest. PhD candidacy qualifying paper, University of Washington, 2014.

[85] John M. Riebold. *The social distribution of a regional change: /æg, ɛg, eg/ in Washington State*. PhD thesis, University of Washington, 2015.

[86] John R. Searle. A classification of illocutionary acts. *Language in society*, 5(01):1–23, 1976.

[87] June E. Shoup. Phonological aspects of speech recognition. In Wayne A. Lea, editor, *Trends in Speech Recognition*, pages 125–138. Prentice-Hall, Englewood Cliffs, 1980.

[88] Freeverse Software. Sound studio (version 3.5.7 for mac), 2008. Newer versions available from Felt Tip Software `http://felttip.com/ss/`.

[89] Swapna Somasundaran and Janyce Wiebe. Recognizing stances in online debates. In *Proceedings of ACL 2009: Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing*, 2009.

[90] Swapna Somasundaran, Janyce Wiebe, Paul Hoffmann, and Diane Litman. Manual annotation of opinion categories in meetings. In *ACL Workshop: Frontiers in Linguistically Annotated Corpora (Coling/ACL 2006)*, 2006.

[91] Pamela Souza, Namita Gehani, Richard Wright, and Daniel McCloy. The advantage of knowing the talker. *Journal of the American Academy of Audiology*, 24(8):689, 2013.

[92] Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3):339–373, 2000.

[93] The Wilderdom Store. Survival scenario exercise: Description of a group dynamics team building exercise. `http://wilderdom.com/games/descriptions/SurvivalScenarios.html`, 2009.

[94] Joshua Tauberer and Keelan Evanini. Intrinsic vowel duration and the post-vocalic voicing effect. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech 2009)*, 2009.

[95] Khiet P. Truong and Dirk Heylen. Disambiguating the functions of conversational sounds with prosody: The case of 'yeah'. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association (Interspeech 2010)*, 2010.

[96] Vincent Q. Vu. ggbiplot: A ggplot2 based biplot. [R package version 0.55], 2011. Available online: `https://github.com/vqv/ggbiplot`.

[97] Michael Ward. Portland dialect study: The fronting of /ow, u, uw/ in Portland, Oregon. Master's thesis, Portland State University, 2003.

[98] Nigel Ward and Wataru Tsukahara. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, 32(8):1177–1207, 2000.

[99] Alicia B. Wassink. Sociolinguistic patterns in Seattle English. *Language Variation and Change*, 27:31–58, 3 2015.

[100] Alicia B. Wassink and Chris Koops. Quantifying and interpreting vowel formant trajectory information, 2013. Paper presented at New Ways of Analyzing Variation (NWAV 42), Pittsburgh.

[101] Alicia B. Wassink and John M. Riebold. Individual variation and linguistic innovation in the American Pacific Northwest, 2013. Paper presented at the Chicago Linguistic Society 49 Workshop on Sound Change Actuation.

[102] Robert Weide. The Carnegie Mellon pronouncing dictionary [CMUdict v. 0.6], 2005. Available online: `http://www.speech.cs.cmu.edu/cgi-bin/cmudict`.

[103] Anne Wichmann. Looking for attitudes in corpora. *Language and Computers*, 40:247–260, 2002.

[104] Janyce Wiebe, Theresa Wilson, and Claire Cardie. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2–3):165–210, 2005.

[105] Theresa Wilson. Annotating subjective content in meetings. In *Proceedings of the Language Resources and Evaluation Conference*, 2008.

[106] Theresa Wilson and Stephan Raaijmakers. Comparing word, character, and phoneme n-grams for subjective utterance recognition. In *Proceedings of the 9th Annual Conference of the International Speech Communication Association (Interspeech 2008)*, 2008.

[107] Richard Wright and Pamela Souza. Comparing identification of standardized and regionally valid vowels. *Journal of Speech, Language, and Hearing Research*, 55:182–193, 2012.

[108] Jiahong Yuan and Mark Liberman. Speaker identification on the SCOTUS corpus. In *Proceedings of Acoustics '08*, 2008.

[109] Victoryia Zayats, Mari Ostendorf, and Hannaneh Hajishirzi. Multi-domain disfluency and repair detection. In *Proceedings of the 15th Annual Conference of the International Speech Communication Association (Interspeech 2014)*, 2014.

Appendix A
DEMOGRAPHIC QUESTIONNAIRE

With both subjects seated in the recording booth, after microphones have been adjusted, the researcher orally discusses the questions on the following page with each subject, notes the responses on the paper form, and later enters them into a secured subject database used in previous and related studies. Uses for a few of the fields are noted here.

The unique speaker ID# is comprised of:

- A two-letter dialect region code, which is always "NW" (Northwest) for ATAROS but may include other regions for other studies in the database
- M or F for sex (male/female)
- A three-digit number indicating the nth male/female subject for the region

For example: "NWF025" indicates the 25th Northwest female in the database.

Channel indicates the speaker's channel in the stereo recording (left/right).

Group indicates the nth dyad recorded for ATAROS (used for internal file organization).

If the recording session is cut short, the completed tasks are indicated in the Tasks field (unused for the ATAROS subjects, who all completed all tasks).

Question 1 ensures that subjects are natives to the Pacific Northwest.

Question 2 gathers coarse impressions of the likely network structure of subjects' home communities. This has been a factor in many sociolinguistic studies, but it is not considered in the current analyses presented here.

Question 3 ensures that subjects are native English speakers and gathers information on their language/dialect experiences, which may affect their own language use. Multilingual subjects were not disqualified as long as they learned English from birth and consistently used it throughout their lives.

ATAROS Demographic Questions Date: ____________________

Investigators: Gina-Anne Levow, PhD (levow@uw.edu), Richard Wright, PhD (rawright@uw.edu), Mari Ostendorf, PhD (ostendor@uw.edu), Valerie Freeman, MA (valerief@uw.edu); Dept. of Linguistics, University of Washington

Sex: M F Birth Year: _____________ ID#: ___________________ Channel: L R

Group: ________ Tasks: **ALL** [Map Inv Sv Cat Bud] Recorded by: __________

1. Where did you live when you were growing up? (incl. ages lived in each place)

2. Would you say that the community was close-knit with everyone knowing each other, or did people have more connections to other communities? For example, did most people work in the community, or somewhere else? Did people have one close group of friends for a long time, or did they have some friends here, some there, some living far away…?

Ages	Place	Network (close or loose + details)

3. What other languages or dialects do you have experience with? (Continue on back if more.)

(1) Language: ____ Known since age: ________

How learned, used: ___________________________________ Proficiency: _____________

(2) Language: ____ Known since age: ________

How learned, used: ___________________________________ Proficiency: _____________

Appendix B
ELICITATION MATERIALS

The following pages present elicitation materials used in the five collaborative tasks in the order in which they are administered.

For the *Map Task*, speakers are seated across from each other at a small table. Each is given a clipboard with one of the “store maps” (pp. 88-89) representing different ways the same household items could be arranged in three aisles of a superstore. The researcher points out the instructions at the top of the maps, orally explains the procedures, and answers speakers’ questions before leaving the booth while the speakers complete the task.

For the *Inventory Task*, speakers stand side by side facing a wall of the recording booth covered in felt on which a few cards have been placed next to lines representing aisles in the “store map” they will fill in. The arrangement of these cards is shown in Figure B.1. The researcher explains the task with oral instructions such as the following:

> You are the co-managers of a new superstore. Your job is to tell your employees where to shelve the new inventory by placing each product on a map of the store. You don’t have time to rearrange things, so once you place an item on the map, you cannot move it. That means you must come to an agreement about where each item belongs before placing it on the map. A few items have already been shelved.

Speakers are given a box of cards backed with Velcro, each printed with the name of a household item, to place on the wall map. These items are the same as those in the Map Task. A photo of an example completed map arrangement is shown in Figure B.2. The researcher answers any questions and leaves the booth while speakers complete the task.

Store Map – Left Speaker

You each have a map of a different super store. Describe your maps to each other to find out how they're different. For example, you could say, "Do you have shoes? They're next to sandals on my map." You can't look at each other's maps until you've discussed all the items.

boating supplies
fish hooks
heavy cable
tow rope
fishing nets
cook stoves
box knives
electric heaters
siding
power cords
five-pound weights
bundles of wood
axes
peat moss
mouse traps
half-inch tubing
matches
saw
bundles of sticks
duct tape
canvas bags
wet suits
cushions

hats
sweaters
coats
vests
boots
socks
jackets
toys
books
travel guides
flags
paper
scissors
chocolate bars
oatmeal
doughnuts
soy beans
beets
cake
dried figs
shoelaces
pet food
canned peas

refrigerator magnets
egg timers
toilet paper
toothpaste
soap
face cream
tweezers
plastic jugs
buckets
paper bags
cups
bottled water
backpacks
sugar
bagels
cake mix
eggs
butter
cookies
ice cream
whiskey
juice

Store Map – Right Speaker

You each have a map of a different super store. Describe your maps to each other to find out how they're different. For example, you could say, "Do you have shoes? They're next to sandals on my map." You can't look at each other's maps until you've discussed all the items.

fish hooks
boating supplies
cushions
tow rope
fishing nets
canvas bags
heavy cable
cook stoves
box knives
electric heaters
power cords
siding
saw
bundles of wood
axes
matches
peat moss
bundles of sticks
duct tape
five-pound weights
mouse traps
half-inch tubing
wet suits

chocolate bars
oatmeal
sugar
eggs
doughnuts
soy beans
butter
cookies
beets
cake
dried figs
ice cream
bagels
cake mix
whiskey
juice
toys
books
travel guides
paper
scissors
pet food
flags

canned peas
refrigerator magnets
egg timers
toilet paper
tweezers
soap
face cream
toothpaste
bottled water
paper bags
buckets
plastic jugs
cups
backpacks
hats
boots
sweaters
coats
vests
shoelaces
socks
jackets

boating supplies	toys	sugar	
	travel guides	refrigerator magnets	ice cream
		cake mix	
siding			
		toilet paper	eggs
	mousetraps		
	pet food	soy beans	
power cords			

Figure B.1: Inventory Task initial arrangement

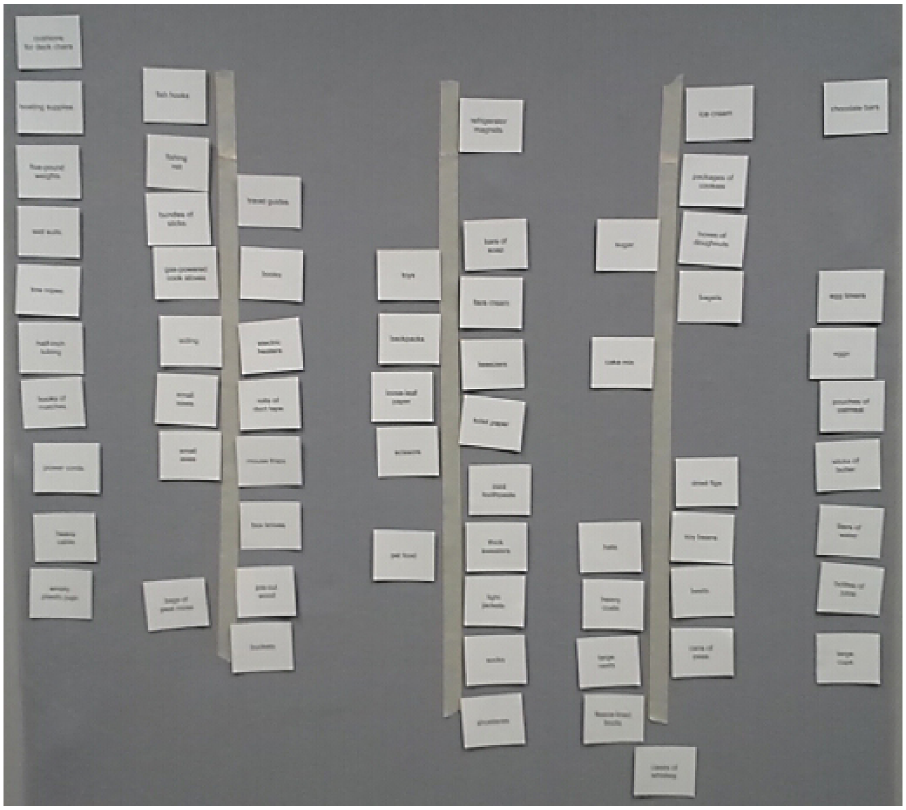
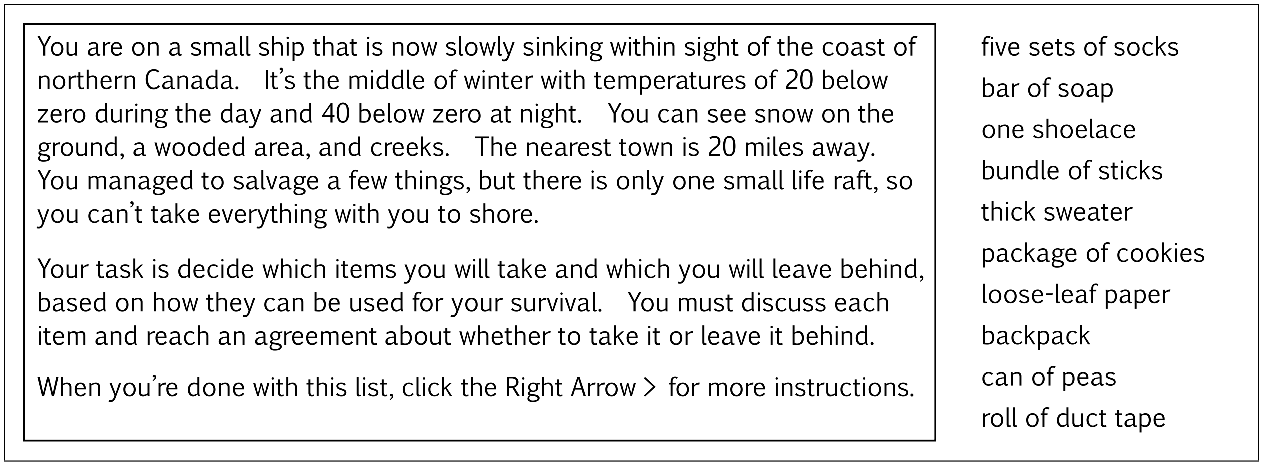


Figure B.2: Completed Inventory Task example

For the *Survival Task*, speakers are seated in front of a computer screen with the instructions shown in Figure B.3. After discussing the items on the initial list, speakers press a key to advance to the next screen, which adds another 10-item list to discuss. This is repeated until all items are visible, as shown in Figure B.4. The items are the same as those in the Map and Inventory Tasks. Before beginning, the researcher reads and explains the instructions orally, demonstrates advancing to the next screen, answers any questions, and then leaves the booth while speakers complete the task. Speakers are told that they can revisit items from previous lists and make their decisions based on any criteria they choose; "the important thing is to discuss each item and make a decision about it." This is encouraged by the gating procedure, which is designed to help speakers focus on a smaller number of items at a time.

The procedures for the *Category Task* are the same as those in the Map Task, with speakers seated across from each other with clipboards, except that the lists (pp. 93-94) are made up of expenses and services that could be funded by a county budget. The researcher goes over the instructions orally, emphasizing that speakers should simply find differences between their lists without making decisions about items to cut, and then leaves the booth while speakers complete the task.

For the *Budget Task*, speakers are seated in front of a computer screen with the instructions and budget items shown in Figure B.5. The items are the same as those in the Category Task. The researcher reads and explains the instructions orally, answers any questions, and then leaves the booth while speakers complete the task. Speakers are told that they can choose how many items to cut from the "department" categories as long as they cut the same number from each, but again, "the important thing is to discuss each item and make a decision about it." The purpose of the restriction is to encourage repeated or extended discussion of multiple items.



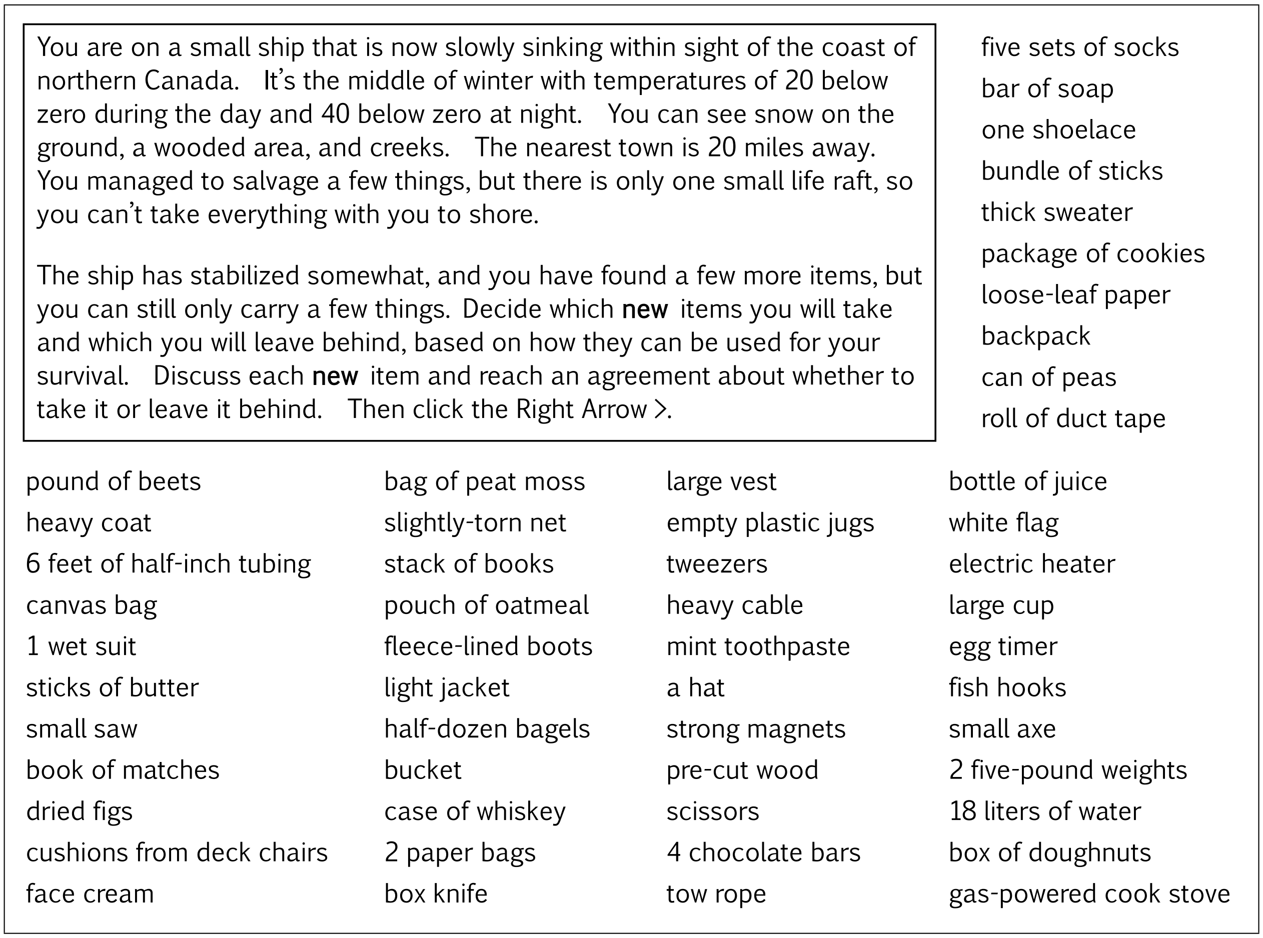
You are on a small ship that is now slowly sinking within sight of the coast of northern Canada. It's the middle of winter with temperatures of 20 below zero during the day and 40 below zero at night. You can see snow on the ground, a wooded area, and creeks. The nearest town is 20 miles away. You managed to salvage a few things, but there is only one small life raft, so you can't take everything with you to shore.

Your task is decide which items you will take and which you will leave behind, based on how they can be used for your survival. You must discuss each item and reach an agreement about whether to take it or leave it behind.

When you're done with this list, click the Right Arrow > for more instructions.

five sets of socks
bar of soap
one shoelace
bundle of sticks
thick sweater
package of cookies
loose-leaf paper
backpack
can of peas
roll of duct tape

Figure B.3: Survival Task initial screen



You are on a small ship that is now slowly sinking within sight of the coast of northern Canada. It's the middle of winter with temperatures of 20 below zero during the day and 40 below zero at night. You can see snow on the ground, a wooded area, and creeks. The nearest town is 20 miles away. You managed to salvage a few things, but there is only one small life raft, so you can't take everything with you to shore.

The ship has stabilized somewhat, and you have found a few more items, but you can still only carry a few things. Decide which **new** items you will take and which you will leave behind, based on how they can be used for your survival. Discuss each **new** item and reach an agreement about whether to take it or leave it behind. Then click the Right Arrow >.

five sets of socks
bar of soap
one shoelace
bundle of sticks
thick sweater
package of cookies
loose-leaf paper
backpack
can of peas
roll of duct tape

pound of beets
heavy coat
6 feet of half-inch tubing
canvas bag
1 wet suit
sticks of butter
small saw
book of matches
dried figs
cushions from deck chairs
face cream

bag of peat moss
slightly-torn net
stack of books
pouch of oatmeal
fleece-lined boots
light jacket
half-dozen bagels
bucket
case of whiskey
2 paper bags
box knife

large vest
empty plastic jugs
tweezers
heavy cable
mint toothpaste
a hat
strong magnets
pre-cut wood
scissors
4 chocolate bars
tow rope

bottle of juice
white flag
electric heater
large cup
egg timer
fish hooks
small axe
2 five-pound weights
18 liters of water
box of doughnuts
gas-powered cook stove

Figure B.4: Survival Task final screen

Budget Categories – Left Speaker

You are on the county committee in charge of balancing the budget. Two advisers have identified areas that could be cut to reduce costs. You each have different lists, and you need to discuss how they're different. For example, you could say, "Do you have traffic lights? They're under Transportation in my list." You can't look at each other's lists until you've discussed all the items.

Transportation

towing services
speed limit signs
additional bus stops
taxi stops
boating licenses

Recreation

junior soccer league
fishing licenses
bookkeeping classes
boys basketball club
football stadium upkeep
cooking classes
hunting tags
football equipment

Public Health

reproductive education
reusable bag campaign
hospital additions
chicken pox vaccinations
STD education
needle exchange
veterinary hospital
egg farm regulations

Infrastructure

pothole maintenance
weed control
subway system
invasive species removal
flag pole repair
public bus upkeep
drainage ditches

Adult Education

job training programs
teaching certificates
acting coaches
massage certificates
math tutors
tattoo artist licenses

Public Safety

sex offender database
stray cat spaying
toxic waste disposal
bagel factory inspections
kitten & puppy adoption
dog catcher
pest control

Low-Income Programs

soup kitchens
community news ads
prenatal check-ups
housing assistance
veterans' medical assistance
food bank
public access station
neighborhood watch support

Education

poetry books
sex ed
custodians
speech therapy
assistant cooks
special ed teachers
sugar-free juice machines
note-takers (disability services)
music teachers

Budget Categories – Right Speaker

You are on the county committee in charge of balancing the budget. Two advisers have identified areas that could be cut to reduce costs. You each have different lists, and you need to discuss how they're different. For example, you could say, "Do you have traffic lights? They're under Transportation in my list." You can't look at each other's lists until you've discussed all the items.

Law Enforcement

towing services
flag pole repair
speed limit signs
boating licenses
fishing licenses
hunting tags
egg farm regulations
neighborhood watch support
sex offender database

Transportation

additional bus stops
pothole maintenance
subway system
taxi stops
public bus upkeep

Outreach Programs

community news ads
food bank
junior soccer league
soup kitchens
housing assistance
boys basketball club
public access station
football stadium upkeep

Education Assistance

speech therapy
math tutors
special ed teachers
music teachers
note-takers (disability services)

Professional Services

teaching certificates
massage certificates
tattoo artist licenses
job training programs
custodians

Public Health

reproductive education
chicken pox vaccinations
STD education
hospital additions
needle exchange
veterans' medical assistance
prenatal check-ups
sugar-free juice machines

High School Programs

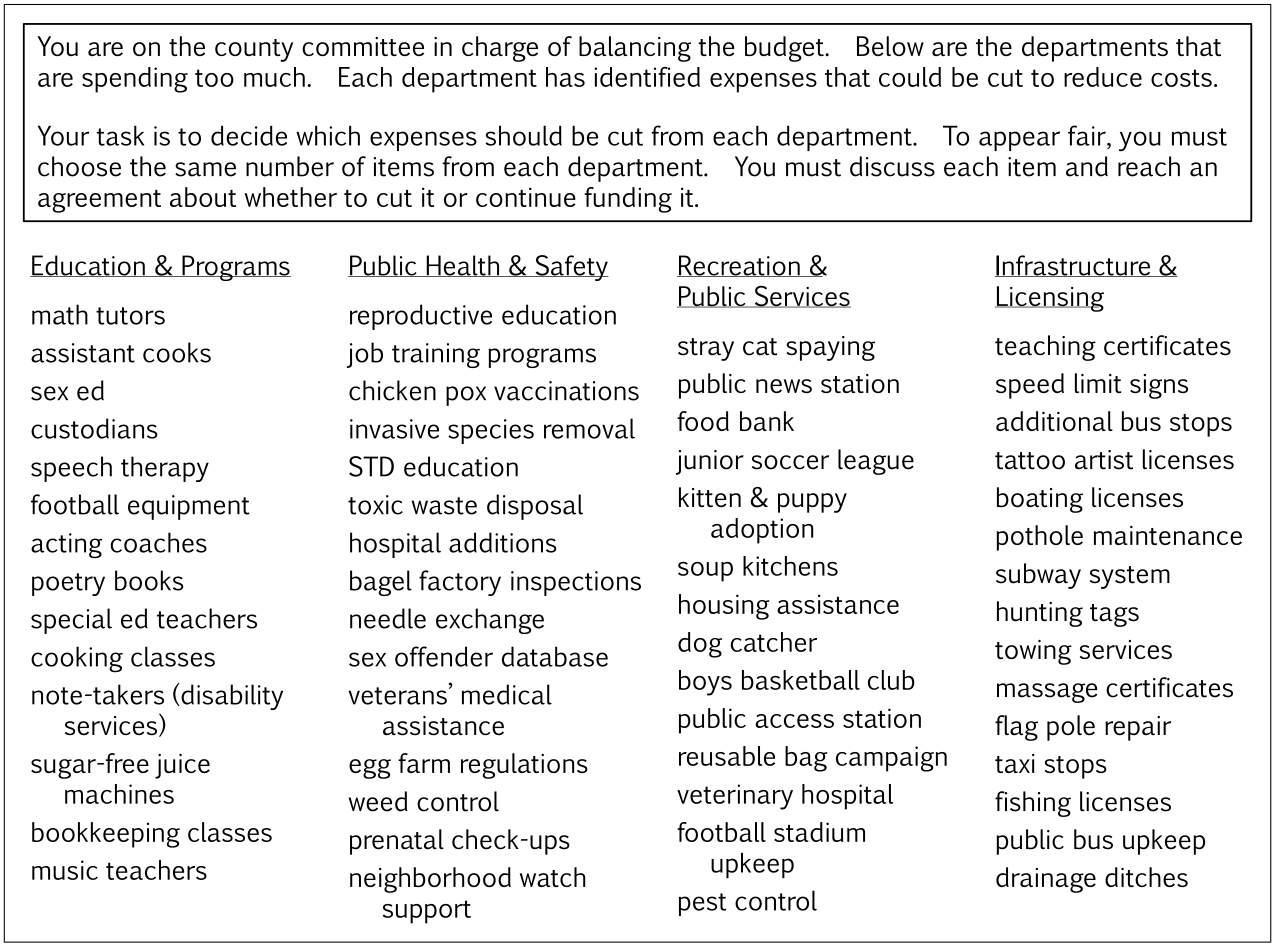
poetry books
acting coaches
cooking classes
sex ed
assistant cooks
bookkeeping classes
football equipment

Animal Control

stray cat spaying
kitten & puppy adoption
dog catcher
veterinary hospital
pest control

Environmental Safety

invasive species removal
reusable bag campaign
drainage ditches
toxic waste disposal
weed control
bagel factory inspections



You are on the county committee in charge of balancing the budget. Below are the departments that are spending too much. Each department has identified expenses that could be cut to reduce costs.

Your task is to decide which expenses should be cut from each department. To appear fair, you must choose the same number of items from each department. You must discuss each item and reach an agreement about whether to cut it or continue funding it.

Education & Programs

math tutors
assistant cooks
sex ed
custodians
speech therapy
football equipment
acting coaches
poetry books
special ed teachers
cooking classes
note-takers (disability services)
sugar-free juice machines
bookkeeping classes
music teachers

Public Health & Safety

reproductive education
job training programs
chicken pox vaccinations
invasive species removal
STD education
toxic waste disposal
hospital additions
bagel factory inspections
needle exchange
sex offender database
veterans' medical assistance
egg farm regulations
weed control
prenatal check-ups
neighborhood watch support

Recreation & Public Services

stray cat spaying
public news station
food bank
junior soccer league
kitten & puppy adoption
soup kitchens
housing assistance
dog catcher
boys basketball club
public access station
reusable bag campaign
veterinary hospital
football stadium upkeep
pest control

Infrastructure & Licensing

teaching certificates
speed limit signs
additional bus stops
tattoo artist licenses
boating licenses
pothole maintenance
subway system
hunting tags
towing services
massage certificates
flag pole repair
taxi stops
fishing licenses
public bus upkeep
drainage ditches

Figure B.5: Budget Task screen

Appendix C
TRANSCRIPTION GUIDE

General instructions

1. Open the stereo sound file for the task in Praat. Extract both channels. (It is easier to understand each speaker if you transcribe them separately.)
2. Create one interval TextGrid for the stereo file, or if an unfinished grid exists, open it, and view it with the sound file for the first speaker/channel. Name the first tier with the speaker code of the left-channel speaker (channel 1). Add a second interval tier and name it with the right-channel speaker code (channel 2). See *file naming* below.)
3. Use boundaries to mark off 'spurts' in each speaker's tier (stretches of speech surrounded by at least 500 ms of non-speech). If there is vocal noise (laugh, cough, etc.; see VOC tags below) which is easily separable from the speech, mark it separately.
4. In every spurt interval, type exactly what the speaker said, word for word, using the conventions below. When finished with the first speaker/channel, view the second speaker/channel sound file with the TextGrid for both speakers.
5. Save the TextGrid with the same file name as the stereo sound file.

File naming: Files are named with the following convention: left speaker ID - right speaker ID - task code. Example: NWF011-NWM052-3I.wav = the 11th Northwest female in the left channel, the 52nd Northwest male in the right channel, completing the Inventory task. (See Appendix A for more on speaker codes.) Task codes:

1D:	Demographics	4S:	Survival
2M:	Map	5C:	Category
3I:	Inventory	6B:	Budget

Conventions[1]

I. Words

a. Transcribe **words** in standard orthography (spelling). Use hyphens in compounds, as found in a dictionary.

b. **Truncated words** (words cut off in the middle by the speaker): mark the cut-off with a hyphen. Do not fill in the whole word, just the part that was said:

> That's not sm-

c. **Numbers**: spell out as words (exactly as they were said) in standard orthography:

> Two thousand and three point six, twenty-five percent, twenty ten.

d. **Pronounceable acronyms**: type using the same capitalization normally used for them in writing:

> UNESCO, MoMA

e. **Spoken letters**: Type each capital letter followed by an underscore:

> Things to discuss. A_, the budget, B_, the new office space.
> the log of X_ plus N_

i. When spoken letters come in clusters, don't put spaces between them:

> His name is Hudson, H_U_D_S_O_N_.
> The C_I_A_ and the F_B_I_ but not FEMA or NASA.

ii. Use a hyphen to combine spoken letters with non-letters:

> C_-three-P_O_ and R_-two-D_-two

f. **Discourse markers**: word forms typical in spoken discourse which may not be in the dictionary but which have conventionalized spellings. See the Transcription Chart for a full list. Examples:

> gonna, wanna, shoulda, coulda, woulda, oughta, sposta, useta, kinda,
> cuz, nah, oh, uh-oh, uh-uh ("no"), uh-huh ("yes"),

[1] Conventions modified from the ICSI Meeting Corpus transcription guidelines [69].

mm-hm, mm (“yes” or vocal nod),

uh, um: the only two for **filled pauses** (um with nasality, uh without)

Note: use comment brackets (described below) to note lengthening, unusual/expressive intonation, pronunciation, etc.

g. **Vocal gestures**: Word-like vocalizations that have a socially-recognized meaning. See Transcription Chart for a list.

h. **“Weird” pronunciation** *for the speaker*: use the PRN tag. Use when speakers’ pronunciation departs from their norms, as with extremely lengthened words/segments or non-words arising from speech errors (these sound like complete words, not truncations). How to PRN-tag:

   i. Put an apostrophe (’) before the stretch of non-canonical pronunciation (a stretch may be more than one word).

   ii. Spell the words in standard orthography.

   iii. Then put the tag (all tags use curly brackets): {PRN description}.

      ’thanks {PRN tanks}

      Go on, ’get! {PRN git}

   iv. Do not use IPA, quotation marks, apostrophes, or any brackets inside { }.

   v. Comments for speech errors can be left blank or described:

      ’posh {PRN error ow}

      ’posh {PRN mispronounced}

   vi. Lengthening (and other embellishments) which are within the normal range of pronunciation variation are noted in QUAL comments (discussed below) instead of PRN comments.

   vii. Do not use IPA symbols, quotation marks or any other brackets < > [ ] inside any transcription or { } tag.

i. **Foreign words**/phrases that aren’t commonly used in English: Mark as with a PRN tag but use an abbreviation for the language instead of PRN, and gloss the meaning, if known:

'Nein! {GER no}

'cum grano salis. {LATIN with a grain of salt}

'tango de la muerte. {SPAN}

j. **Uncertainty (you're not sure what was said)**: Use parentheses around the uncertain portion:

   i. If you're reasonably sure, put the words in parentheses: (word or phrase)

   ii. If you can reasonably identify the number of syllables but not the words, put the number and x in parentheses, e.g., (3x) for a three-syllable utterance

   iii. If totally indecipherable: (??)

II. Utterances: Similar to sentences or phrases, begin with capital letter and end with punctuation or a comment tag. **Punctuation**: Pay attention to the pragmatic force: what does the speaker mean to do with the utterance? Don't rely on just syntax or just intonation alone (e.g. some word-orders look like statements but are questions; rising intonation doesn't always indicate a question).

a. **Exclamations**: end with an exclamation point: Awesome!

b. **Statements**: end with a period.

c. **Questions**: end with a question mark. Questions with declarative syntactic forms also end with question marks (they look like statements but are intended as questions): And you're a student? (meaning "are you a student?"). Rising intonation used to elicit feedback (as if to say, "know what I mean?" or "are you following me?") does not use question marks.

d. **Disfluencies and incomplete utterances**: end with a space and a hyphen unless the last word is truncated (then, use the truncation hyphen with no space):

   It'd be nice, but - but I - I do- I don't wanna count on it.

e. If the incomplete utterance was clearly a question, the hyphen can be followed by a space and a question mark (- ?):

   So, uh, what was the date? Monday or - ?

f. **Emphasis**: When a word is strongly emphasized (it stands out and has more emphasis than expected if reading the words alone), put an asterisk directly before the word (no space):

> So you weren't really *exposed or anything, right?

g. **Intonation and voice quality** that stands out as strong or unusual: Use a QUAL comment tag to describe:

> Face cream? {QUAL incredulous intonation}
>
> Books. {QUAL lengthened}
>
> The new X_Men movie, {QUAL rising intonation} .. totally sucked. {QUAL creaky}

h. Commas: use as in standard orthography.

i. Do not use quotation marks, colons, or semicolons anywhere.

III. **Vocalizations** (not words, possibly metalinguistically meaningful, but not necessarily): use a VOC comment tag to describe.

a. Use only these five:

> {VOC laugh}
>
> {VOC cough}
>
> {VOC breath} (for loud breaths, e.g., signaling start/end of turn)
>
> {VOC start} (hesitation noise that is clearly attempting speech)
>
> {VOC mouthnoise} (everything else)

b. When these are adjacent to speech but easily separable, mark them in intervals separate from speech.

c. Use a separate tag for each vocalization, even in succession:

> {VOC laugh} {VOC cough}

d. If the sound overlaps speech, use a QUAL tag instead:

> Right! {QUAL laughing}

IV. **Non-vocalized sounds**: use a NVC comment tag, with optional description.

   a. Use a separate tag for each noise, even if in succession:

      {NVC door slam} {NVC}

   b. In general, mark only when they occur during a speaker's turn. When possible, mark in intervals separate from speech.

V. **Silence/Pause**:

   a. use two periods with a space on either side for pauses within a spurt (shorter than 500 ms):

      Um .. yeah. Oh - .. Right.

   b. Mark empty intervals with {SIL} (every interval without speech):

      - When finished transcribing intervals with speech, select the TextGrid in the Objects window > Modify > Modify interval tier > Replace interval text.
      - Enter the tier number (e.g. first do tier 1, then do this for tier 2).
      - For Interval Range, use 0 in the first box and the highest interval for that tier (when viewing the TextGrid, the number below the tier name on the far right of the tier).
      - Make the Search box blank.
      - Enter {SIL} in the Replace box.
      - Leave Literals checked. Hit OK.
      - Repeat for each tier.

TRANSCRIPTION CHART

A quick distillation of the guidelines[2]. Keep this chart handy when transcribing.

Tags

* = Place an apostrophe at the beginning of the speech with this feature.

Do not use IPA symbols, double-quotes, slashes, brackets (“ ” / \[ ] < >) inside { } tags.

Non-Vocalized Noises	{NVC <optional description>}
	{NVC} {NVC door slam}
Silence	
less than 500 ms	.. (space, two periods, space)
500 ms or more	{SIL} (the only thing in a silent interval)
Vocalizations	use only these five:
laughter	{VOC laugh}
coughing	{VOC cough}
breathing (loud and/or meaningful)	{VOC breath}
false start/hesitation noise	{VOC start}
any other mouth noise (sniff, smack, etc.)	{VOC mouthnoise}
Quality	{QUAL <description>}
	{QUAL laughing}
	{QUAL lengthened}
***Pronunciation**	'<word or phrase> {PRN <description>}
	'Exterminate {PRN Dalek voice}
	'Yes {PRN yesh}

[2] Special thanks to Heather Morrison for the original compilation of this chart.

***Foreign words**	{<LANG> <word/phrase, if known>}
	'Nein! {GER no}
	'Ciamar a tha thu? {GAE}

Uncertainty	(can't tell what's said)
Reasonably sure	(word or phrase in parentheses)
Syllables but not words identifiable	(<number of syllables>x): (3x)
Completely indecipherable	(??)

Punctuation

Capitalization	(standard) Mary and I live in Seattle.
Truncated words	<word bit>- (no space, dash) Well, b- but, I like dogs and ca-
Statements	<Sentence>. (period) This is a sentence.
Exclamations	<Sentence>! (exclamation point) Woo hoo!
Questions	<Sentence>? (question mark) What is your name? That is your quest? It's your favorite color, right?
Disfluencies/Incomplete Sentences	<phrase> - (space, dash, space) Well, he was - What class was that? Physics or - ?
Emphasis	* <word> You aren't *really the Dayman.

Filled Pauses

With nasality	um
No nasality	uh

Spelling

Words	standard spelling
Numbers	spelled out as said forty-two a hundred and fifty percent
Pronounceable acronyms	standard capitalization UNESCO, FEMA, NASA
Spoken letters	<capital letter>_ (underscore) the letter A_ C_-three-P_O_ F_B_I_, C_I_A_

Discourse markers, vocal gestures, variable/non-standard spellings

For these pseudo-words, use only these spelling variants. When encountering a new one, discuss with the transcription supervisor to determine the spelling to add to this list.

alright	okay	kay	mm-kay	mm-hm
hm	mm	uh-huh (‘yes’)	uh-uh (‘no’)	uh-oh
oh	ah	aha	nah	psh (scoff)
dunno	gonna	wanna	lotta	outta
woulda	coulda	shoulda	oughta	sposta
useta	hafta	kinda	sorta	cuz (‘because’)

Appendix D
STANCE STRENGTH/POLARITY ANNOTATION GUIDE

Annotate task sound files that have been transcribed and time-aligned. Note: This process is also abbreviated "coarse annotation."

1. Open the stereo sound file in Praat. Make sure the audio is loud enough to clearly understand the speech of both speakers. (If it isn't: from the Objects window, Modify > Scale peak, enter a number up to 0.99.)
2. Open and View the transcribed TextGrid in Praat (the aligned tiers are not needed). Using the Tier menu, duplicate each speaker tier; place the duplicate below the original, name it 'coarse,' and remove all text from the duplicate, so only the boundaries remain.
3. Mark each spurt (interval with transcribed speech) for stance strength and polarity using the labeling conventions below. Listen to both speakers at once, but annotate one speaker all the way through the task, and then listen again to annotate the other.
4. Save the TextGrid with '-coarse-' and your initials added to the end of the file name, e.g., NWF001-NWM022-6B-coarse-VF.TextGrid
5. **Second pass annotation** (done by a second annotator): Follow the annotation procedures, checking/correcting all labels and removing all asterisks (*). Save the TextGrid without initials in the file name, e.g., NWF001-NWM022-6B-coarse.TextGrid

Stance presence/strength

Mark each spurt with one of the following. Add * after the number if uncertain.

Label	Description
0	**no stance** • reading: "I have shoes, socks, jackets..."

	• factual questions/answers: “Did we see socks before?” “Yeah, I have them.” • conversation managers: “Okay, next item.” • backchannels (I acknowledge you spoke, keep talking.)
1	**weak** • mild/cursory agreements: “Sure”, “yeah”, “okay/good/fine” • mild opinion: “that’s good,” “we should keep that” • mild praise/encouragement: “good idea”, “that makes sense” • offer solution: “Should we put it here?” • solicit opinion: “What do you think?” • facts/reasons, personal credibility as support, without strong feeling: “Let’s put it here because these are alike.” “Yeah, I’ve seen it done that way.”
2	**moderate** • stronger versions of items under 1 • question other’s opinion: “Why?” “Do you really think so?” • offer alternative: “Or how about here instead?” • mild-moderate disagreement; can be hedging: “Well/maybe {lengthened}, I don’t know” or confident but not emotional: “I disagree”, “No, let’s not.”
3	**strong** • emphatic, excited, strong versions of items under 2 • loaded and/or emotional expressions or exclamations
x	**unclear; activated but not “stancey”** • truncations (can’t tell what they’re trying to say or do) • “Oh! That’s what that means!” (sounds stancey but isn’t) • comments on the task/items that aren’t taking an identifiable stance, e.g., laughing at/mocking an item: “Face cream? {incredulous}”

Polarity

For spurts with stance (not labeled 0 above), add one of the following to the stance strength number. Add * after the symbol if uncertain.

Label	Description
+	**positive** • agreement: “yes, yeah, sure, okay, fine” • supportive reasons/comments; encouragement: “Good idea. That makes sense” • intonation that conveys positivity
-	**negative** • disagreement, contradiction, reasons against, hedges: “Well... {lengthened}” • question other’s opinion • intonation that conveys negativity
(none)	**neutral** • Neither positive nor negative
x	**unclear** • Can’t tell; could be positive, negative, or neutral

Possible strength + polarity label combinations

	Polarity			
Strength	neutral	positive	negative	unclear
none	0			
weak	1	1+	1-	1x
moderate	2	2+	2-	2x
strong	3	3+	3-	3x
unclear	x			

Appendix E
STANCE TYPE ANNOTATION GUIDE

Annotate task sound files that have been transcribed and time-aligned (and optionally annotated for strength/polarity). Note: This process is also abbreviated "fine annotation."

First pass

1. Open the stereo sound file in Praat. Make sure the audio is loud enough to clearly understand the speech of both speakers. (If it isn't: from the Objects window, Modify > Scale peak, enter a number up to 0.99.)
2. Open the aligned or coarse-annotated (for strength/polarity) TextGrid in Praat. (If the files have been broken by speaker, open both single-speaker TextGrids.) Add a tier for each speaker called 'fine' with your initials (e.g., 'fine VF').
3. Mark off and annotate stance acts following the type guide below. Copy boundaries exactly from the 'word' tier produced by the forced-aligner. Listen to both speakers at once, but annotate one speaker all the way through the task, and then listen again to annotate the other.
4. Save the TextGrid(s) with '-fine-' and your initials added to the end of the file name, e.g., NWF001-NWM022-6B-fine-VF.TextGrid. If you used single-speaker TextGrids, merge them back together before saving, so all available tiers (phone, word, transcription, (coarse), fine) for channel 1/left speaker are first, followed by all tiers for channel 2/right speaker.

Second pass

To be done after one person has fully type-annotated a file.

1. As above, open the stereo sound file and the first-pass annotated TextGrid in Praat. (If the files have been broken by speaker, open both single-speaker TextGrids.) Duplicate each fine annotation tier; place the duplicate below the original, change to your initials (e.g., 'fine HM'), and remove all text from the duplicate, so only the boundaries remain.
2. Listen to the stereo audio file while checking the annotations (first for one speaker, and then listen again while checking the second speaker). Put labels in your new tier only when you disagree with the original, or when the first annotator added a * (asking for second opinion). Change boundaries only in your tier when you disagree with splits/locations made by the first annotator. Add a * in your tier when you want a third opinion.
3. Save the TextGrid with your initials added to the end of the file name, e.g., NWF001-NWM022-6B-fine-VF-HM.TextGrid. If you used single-speaker TextGrids, merge them back together before saving, so all available tiers (phone, word, transcription, (coarse), fine) for channel 1/left speaker are first, followed by all tiers for channel 2/right speaker.

Finalization (third pass)

To be done after two people have fully type-annotated a file.

1. Follow the second-pass procedures, with the following changes. Duplicate the first annotator's tier, name it 'fine' without any initials, and don't remove the text. Listen to the whole sound file, but only check/correct labels in intervals the second annotator marked. When finished, remove the first and second fine annotation tiers, so yours are the only remaining fine annotation tiers.
2. Save the TextGrid with no initials on the file name, e.g., NWF001-NWM022-6B-fine.TextGrid

General labeling instructions

- Mark boundaries around words/phrases (smaller or larger than a spurt) that serve the following functions and label accordingly. In deciding where to make boundaries:
  - Mark each “stance act,” the chunk of speech that is performing the act(s) below
  - Boundaries may contain pauses, combine or break up spurts.
- Use lowercase letters, and don’t use spaces between letters/symbols.
- When more than one label is needed, order doesn’t matter.

Labeling scheme

Label	Description
o	**Offer opinion**, solution, suggestion, recommendation, what should be done • “(I think) we should..., I say we..., Let’s..., Here’s my idea..., I’m going to...” • “I think, In my opinion, It’s clear to me, I feel like, I (don’t) want” • May include evaluative comments, modifiers, descriptors, or intensifiers (adjectives, adverbs, intensifiers, comparatives) that show bias, opinion, emotion: “obviously, That’s... helpful, useful, important, really good, totally useless.” • May double with Solicit, even only via questioning intonation (label: os or so)
s	**Solicit** opinion, knowledge, agreement, approval, acceptance of opinion/solution • “What do you think?, Is that okay?, Right?, Do you agree?” • Often doubles with Offer: “What if we...?”, “How about...?” or via questioning intonation (label: os or so) • Only mark knowledge-seeking questions if used as a solicitation of expertise/experience and/or if the answer provides support for an opinion, e.g.: A: “We need to decide what kind of county this is.” (label: o) B: “Well, what county do you live in?” (label: s) A: “King County.” B: “Then let’s make it like King County.” (label: co)

c	**Convincing/credibility**: Offer reasons/support for (own/other's) opinion/solution: • Facts/Certainty: "actually, in fact" • Reasons/Explanation: "Because..., That's why..., So..., If we assume..." Note: In reasoning, the "so/because" part may be implied: "This is a baking aisle" (implied support for: "so that's where cake mix belongs") • Experts/Experience/Examples: "I read online, It's in the dictionary, When I was a Boy Scout..., My son..., But I don't know if they have that anymore" • If supporting in agreement with (usually other's) opinion, label: ca or ac (eg as if to say "and here's another reason why that's right") • If used as counter-evidence in disagreement with (usually other's) opinion, label: cd or dc (eg as if to say "and here's another reason why that's wrong / we shouldn't do that")
a	**Agree/Accept/Approve/Affirm**: opinion/solution (other's or own): • "Okay, Sure, Alright, That's fine." • "Yes, I agree, Right, Exactly, Absolutely, That makes sense" • Echo previous (usually other's) opinion, or or complete other's sentence as if to agree/join in the assertion • Accepting/confirming an agreed-upon solution (topic-ender): "Good, Okay, Cool". This may also cover (re-)listing items previously discussed or decided upon: "So we're cutting x, y, z." (They're confirming prior decisions, opening up the chance to confirm or change decisions.)
d	**Disagree/Reject/Contradict** opinion/solution (other's or own): • "I disagree, Absolutely not, No, I don't think so, Hold on, Wait a minute!" • "Well..., Yeah but..." (if followed immediately by explanation, label: dc or cd)

	• Questioning (usually other's) opinion/solution: "Really?" This ay include a reason for disagreeing: "That's redundant" (label: dc or cd) or a solicitation: "Are you sure?" (label: ds or sd) • Could combine with offering an alternative (label: do or od), likely with only intonation indicating the disagreement – use only when the intonation clearly indicates disagreement stronger than polite reluctance (label: ro or or)
r	**Reluctance to accept** (usually) other's opinion/solution: • "Maybe, Well..., Uh..., If you think so, Yeah lengthened..." • Often lengthened, with a pause showing hesitation, and/or reluctant intonation • Often followed by an alternative solution (label: o), or an alternative solution may include reluctant pauses or intonation (label: ro or or) • May accept with reservations: "Okay, I guess" (label: ra or ar)
f	**Soften/Downplay** own opinion/solution; **Hesitate** to offer; uncertain of self: • "Just kidding, I don't know" (may be lengthened/quiet), "But that's just me" • Often said with laughing or sarcasm • Hesitation noises, lengthening; often before giving opinion • Can combine with soliciting opinion, e.g.: "Maybe?" (downplaying own previous solution and asking for approval) (label: fs or sf)
e	**Encourage/Praise** self, speaker, solution/decision (clear "pat on the back"): • "Good idea, Perfect, Nice one!, Good job us, Now we're getting somewhere!" • Only use when clearly encouraging (if confusable with agree/accept, don't use) • If encouraging/praising intonation is strong on another function, add this label
t	**Teamwork/Rapport/Solidarity**: clearly building rapport: • Commenting on the item/task: "this is hard," "pre-cut wood?" *(How funny!)* • Jokes, teasing, sarcasm, commiserating *("We're in the same boat")* • Only use when clear (if you're not sure whether to use it, don't)

i	**Intonation**: something strongly stancey/opinionated: • Use when intonation stands out • “What?!” *(Are you crazy?)*, incredulous, skeptical, mocking, etc. (you don’t have to be able to label the emotion/opinion) • May be used when something is clearly stancey but doesn’t fit into another category or is hard to identify • May be overlaid on another function; add this label to another if it adds to or changes the meaning (e.g. don’t add “i” to an “e” if the only intonational meaning is encouragement)
x	**Can’t tell**: but seems stancey: • Use when stance presence seems obvious, but you can’t identify the meaning • May apply to spurts coarse-labeled x (for unclear stance strength), but be sure to mark off only the words that seem stancey (not necessarily the whole x-labeled spurt)
b	**Backchannel** (verbal nod, I acknowledge you spoke, keep talking) • “Yeah, mm-hm, okay, right” • These are generally strength-labeled as 0 (no-stance)
*	**Unsure** • Add after any label you’re unsure of, to mark for review by another annotator
#	**Alignment problem** • Add after the last label in the interval (no spaces) to indicate that the word boundaries in the interval need to be realigned. • If boundaries for words you need to separate for fine annotation are way off from the audio, mark boundaries in the fine tier to match the audio. If this creates other fine intervals that don’t match the words in the word tier, be sure to mark them with # as well, even if not marking it with any other fine label.