Change in Women's Uptalk Scaling Depending on Conversational Context

Evelyn TeSelle
December 17, 2021

Abstract

This study aims to investigate whether conversational urgency affects the scaling of women's uptalk. Twenty-five woman-identifying participants were placed into 3 experimental groups–L, M, and H–which created low, medium, and high urgency conditions for speaking. Each participant completed a speech task on audio recording. The recordings were subsequently segmented and the scaling of each instance of uptalk was measured. The scaling is the maximum F0 minus the F0 at the beginning of the uptalk rise, representing the amount of pitch rise. Contrary to my hypothesis, the results showed that the L group had the highest mean scaling, the H group had the lowest mean scaling, and the M group was in the middle. This indicates that the participants completing the speech task alone in a room used more extreme uptalk pitch rises than those completing their task in pairs and engaging in back-and-forth conversation.

1 Introduction

"Uptalk" is a term that refers to the use of rising intonation at the end of a declarative utterance. This phenomenon is also sometimes called "high rising terminal." Rising intonation at the end of an utterance is common and unmarked in instances like asking yes-no questions, but it is often marked and even scrutinized when employed on a statement. The use of uptalk is commonly associated with young women and sometimes thought to sound unconfident, vapid, or immature. But there are actually many discourse functions of uptalk, such as confirmation requests and floor holding which serve to establish the flow of backchanneling and turn-taking in an interactive conversation. Confirmation requests are a way of asking the listener if they are following what the speaker is saying. Floor holding is any method of showing the listener that the speaker is still talking–an example of this is forward reference, which is an intonational pattern that directs attention to something the speaker is about to say. There is significant research on these discourse functions of uptalk which has established this feature of "women's speech" as a useful conversational feature.

Although uptalk is used frequently in conversation by all genders, there is significant evidence that women use uptalk more than men. Guy & Horvath (1986) found that in Australian English, women used uptalk in over twice as many instances as the men in their study. Similarly, Ritchart & Arvaniti (2014) found in their study that uptalk comprised 42% of the women's utterances and 20% of the men's utterances. They also found that women employ more extreme pitch contours than men. The evidence that uptalk is used more by women, the association uptalk has with women's speech, and the negative stigma associated with women users of uptalk are all reasons why this study focuses specifically on the discourse functions of uptalk as an aspect of women's speech.

1.1 Research Question

In this study, I investigated if pitch contours of uptalk change depending on conversational urgency. I specifically investigated this phenomenon in women's speech, as women have been shown to employ more extreme pitch contours and to use uptalk more frequently than men. By answering the question of how pitch contours change depending on how a woman needs to use uptalk for a specific function, this study could help to solidify uptalk as an important tool for turn taking, attention holding, and preventing interruption for women speakers.

Specifically, my hypothesis is that the scaling (difference between the F0 at the beginning of the rise and the F0 at the highest point) will be greater in stressful interactive conversations where confirmation requests and floor holding are more urgently necessary than in calm interactions and non-interactive speech. If this hypothesis were to be true, it would further solidify uptalk's purpose as a part of interactive speech, and would show how and when it is employed (and to what degree it is employed) by women when they need to hold the floor, elicit confirmation requests, or establish the flow of turn-taking in conversation. This would serve as a contribution to the field in that it would illuminate an important discourse function for a stigmatized element of women's speech.

2 Literature Review

2.1 Discourse Functions of Uptalk

Guy & Horvath (1986) provide significant background information about the discourse functions of uptalk that is useful for understanding the goals of this study. Their paper begins with an introduction to the "high rising terminal contour" on declarative sentences in Australian English, which at the time was an emergent new language phenomenon (Guy & Horvath, 25). This paper is chiefly useful because of how the researchers detail the discourse functions of the high rising terminal contour, and they are careful to stress that the purpose of the pitch rise is interactional (Guy & Horvath, 25). They write that one of the main discourse functions of the final pitch rise is to ask the listener, "do you understand me?", which can also be referred to as a confirmation request (Guy & Horvath, 26). These confirmation requests also elicit back-channeling, further emphasizing that the pitch rise is an interactional tool (Guy & Horvath, 26). These findings about the discourse function of the high rising terminal contour on declarative sentences lays the groundwork for the present study, as I investigated how different levels of interaction and conversational urgency affect pitch contours. Knowing that pitch contours exist as an interactional tool first and foremost supports the hypothesis that a change in the interactional context and urgency may change the pitch contour.

Banuazizi & Creswell (1999) provide useful background research in their paper which is relevant to the discourse function of uptalk. In their introductory sections, they write that when an utterance ends in a high boundary tone, the speaker is conveying a "forward reference" (Banuazizi & Creswell, 1). This means that the speaker is drawing attention to something that is to come in their next utterance. In a conversation where it is urgent for a speaker to get their point across, a speaker may use this "forward reference" to hold the floor, which supports the hypothesis that speakers may use more extreme uptalk pitch contours in more urgent conversations.

Banuazizi & Creswell (1999) is also relevant to the present study as the main point of their study is investigating how pitch contours change depending on the discourse function of the pitch rise. Specifically, they investigate how the genuine-ness of the question affects how extreme the pitch rise is. The researchers used a corpus, searched for yes-no questions, and coded their final pitch contours as either falls or rises (Banuazizi & Creswell, 6). They then coded the questions for genuine-ness, which involved criteria such as whether or not the speaker already knows the answer to the question and whether they expect a response from the listener (Banuazizi & Creswell, 7). After analysis, the researchers found that a question is more likely to end in a rise if it is a genuine question (Banuazizi & Creswell, 7). These findings are relevant to the current study, because they show that the likelihood and degree of pitch rises changes depending on the discourse function.

Similarly, Fletcher, Stirling, Mushin, and & Wales (2002) paper also offers valuable background insight on the "forward reference" function of uptalk. Fletcher et al. analyzed map task dialogues of speakers of Australian English to find how they used pitch contours in context (Fletcher et al. 229). They found that two types of simple high rises were both used for "forward looking" information, which draws attention to something the speaker is going to say (Fletcher et al. 247). They also found that when speakers took on the "instruction giver" role in the map task, they used lower onset rises to indicate "backward looking" or the end of a turn (Fletcher et al. 247). This information about the discourse function of pitch rises and about how pitch varies depending on the intended function is relevant to the goals of the current study, as it informed my understanding of change in pitch during conversational contexts.

Ritchart & Arvaniti's (2014) *The Form and Use of Uptalk in Southern Californian English* examines how men and women use uptalk differently in terms of discourse function and pitch contours. This study is useful and relevant to the current study, and it lays the groundwork for the concept that uptalk pitch contours can change with discourse function. This study's methods, measurements, and findings all serve as models which I expanded on in my research. Ritchart & Arvaniti (2014) define "scaling" as a measurement of pitch contours in their study, which I used as a measurement of pitch contours in my study. They define scaling as the difference between the F0 at the beginning of the rise and the F0 at the highest point, measured in Hz (Ritchart & Arvaniti, 332).

Most importantly, Ritchart & Arvaniti (2014) also investigate how uptalk pitch rises can become more extreme based on discourse function. The researchers measured the scaling of each instance of uptalk for each of their four predetermined discourse functions (question, confirmation request, floor holding, statement). The results of these measurements can be seen in Figure 1 below (Ritchart & Arvaniti, 333). Note that in the graph, the scaling measurements were converted from Hz to the equivalent rectangular bandwidth (ERB) scale to better compare male

and female voices. The ERB scale gives an approximation to the bandwidths of the filters in human hearing, and is useful for comparisons of this type.

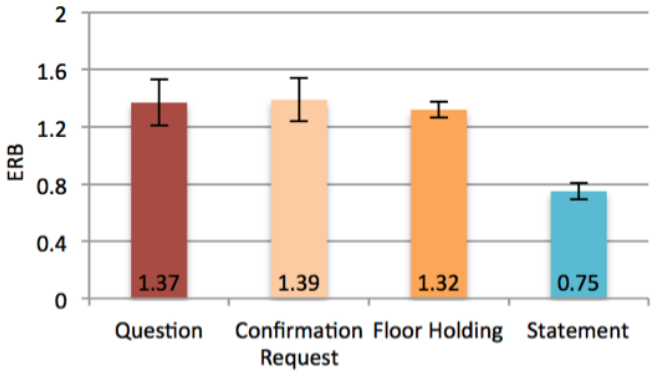

Figure 1: Mean scaling of rises (with standard error bars) per discourse function (Ritchart & Arvaniti, 333)

As we can see from the chart, the degree of the pitch rise does change depending on discourse function. Most interestingly, confirmation requests and floor holding have much more extreme pitch contours than statements (Ritchart & Arvaniti, 333). This is crucial background information for my study, as it provides evidence that more "urgent" functions of uptalk may require more extreme pitch contours.

2.2 Uptalk Users

2.2.1 Uptalk and Gender Identity

Guy & Horvath (1986) provide relevant demographic information about the users of uptalk. The main point of their study was to investigate who uses uptalk the most in order to determine how and why it was gaining popularity in Australia in 1986. Using a corpus, they found that teenagers had a .74 probability of using AQI (Australian Question Intonation) on a declarative sentence, while adults had a .26 probability of using it (Guy & Horvath, 36). Furthermore, they found that men had a .41 probability of using AQI on a declarative sentence, while women had a .59 probability of using it (Guy & Horvath, 39). The present study is focusing on young women's usage of uptalk, and Guy & Horvath's (1986) findings support this focus by showing that both women and young people use uptalk the most. Though their demographic results are from Australia in 1986, they show that young women may have been the pioneers of uptalk which could be one of the reasons it is so heavily associated with them. If women are the creators and primary users of uptalk, it is relevant to study more precisely how they use it.

In addition to providing important groundwork about change in uptalk depending on discourse function, Ritchart & Arvaniti (2014) also validate the common claim that women use uptalk more than men, which is one of the foundational reasons that my study focuses on women's speech only. Overall, the women used uptalk more frequently than the men in the study, with uptalk comprising 42% of the women's utterances and 20% of the men's utterances (Ritchart & Arvaniti, 332). This research and their methods provide a model for my own study and validates the foundational claims for studying uptalk in women.

Linneman (2013) conducted a study on uptalk usage on the show *Jeopardy!* in which he surveyed which participants used uptalk the most and in which contexts. Linneman (2013) coded 100 episodes from 2009 and 2010 as well as 49 special episodes (Linneman, 89). Linneman (2013) found that women on the show used uptalk significantly more than men: when giving a correct response, a man's probability of using uptalk was 27% while a woman's probablity of using uptalk was 48% (Linneman, 94). For incorrect reponses, the probability of using uptalk was 57% for men and 76% for women (Linneman, 94). Finally, Linneman also examined each contestant as uptalk users individually, and found that nine women used uptalk almost exclusively, meaning they uptalked on at least 90% of their utterances, while no men were exclusive uptalk users (Linneman, 94). Linneman notes that while responses in *Jeopardy!* are given in the form of a question, they are still referred to as "answers" and therefore when rising intonation is employed on them it can be classified as uptalk (Linneman, 85).

Tyler (2015) conducted survey study in which they assessed the perceptions of uptalk users. In their first study, Tyler (2015) distributed a survey on Amazon Mechanical Turk (AMT) which contained two audio clips of young women using uptalk (Tyler, 289). The survey asked for general opinions on the speakers use of rising intonation, and responses ranged from positive to neutral to negative (Tyler, 293). Some positive responses described the use of uptalk as exciting or more clearly understandable, while the negative reactions described the audio clips as annoying and lacking confidence (Tyler, 293). In another study, Tyler (2015) distributed audio files of women uptalking on AMT and had the participants rate their agreement with pre-written perceptions of uptalk which were devised using the results from his first study (Tyler, 294). Overall, Tyler (2015) found that listeners perceived uptalk as a way of seeking listener approval, which sounded unconfident and unassertive to some (Tyler, 304). They also found that some listeners thought uptalk sounded polite (Tyler, 304). Finally, Tyler (2015) found that sentences with were uptalked were likely to be marked as "unfinished," which suggests that the listeners may have been picking up on the "forward reference" function of uptalk (Tyler, 304). These findings not only show how uptalk is often perceived negatively when used by women, but they also show that it possesses useful discourse functions like forward reference.

While there is significant research and evidence that women use uptalk more than men, all of these studies classify gender in a narrow and binary manner. It is well established that nonbinary genders exist outside of the classifications of "man" and "woman," which raises the question of how nonbinary people use uptalk. Shar (2018) measured intonational phrases (IPs) as used by various assigned female at birth (AFAB) people of varied gender identities and expressions (Shar, 18). Participants' gender identities and expressions were diverse, with some being cis, trans, and genderqueer and some aligning more masculine, neutral, or feminine (Shar, 19). Shar (2018) organized participants based on survey responses into three categories: masculine aligning, neutral aligning, and neutral to feminine aligning (Shar, 19). Shar (2018) had each participant complete a speech task on recording, and identified and analyzed their pitch contours in Praat according to predetermined guidelines (Shar, 21). After analysis, Shar (2018) found that masculine aligning participants used the most low falling intonation, which may be a method of indexing masculinity with intonation (Shar, 31). It was also found, contrary to expectation, that neutral aligning participants used the most high rising intontation while neutral to feminine aligning participants used less high rising intonation (Shar, 32-32). Shar (2018) speculates that this could be because the neutral aligning participants presented more masculine, and therefore used intonation to index femininity to achieve a form of gender neutrality, while the neutral to feminine aligning participants presented more feminine and therefore felt less motivation to index femininity with intonation (Shar, 32). While this study does not specifically address uptalk but rather intonation in general, it does provide an interesting basis for the idea that speakers may use intonation to index gender, including nonbinary and gender-nonconforming speakers.

Similarly, Schmid & Bradley (2019) conducted a study on the use of pitch on a group of participants with an array of gender identities, including cis women, cis men, trans women, trans men, and nonbinary people who identified as agender, genderqueer, and genderfluid (Schmid & Bradley, 2686). The study had 30 participants total, and each participant was recorded reading a passage aloud in a quiet room (Schmid & Bradley, 2686). The researchers analyzed the recordings for each participant in Praat, taking an average F0 (fundamental frequency) for each sentence and using those averages to calculate an average F0 for each speaker (Schmid & Bradley, 2686). Schmid & Bradley (2019) found that women had the highest average pitch, and men had the lowest average pitch (Schmid & Bradley, 2687). Additionally, the nonbinary participants performed pitch differently from both the men and the women, with an average pitch (as a group) 50 Hz lower than the women and 25 Hz higher than the men (Schmid & Bradley, 2688). The researchers also found that there was not a statistically significant difference between the average pitches of the AFAB and AMAB nonbinary participants, which suggests that assigned gender at birth is not the driving factor in pitch production but rather the indexing of gender instead (Schmid & Bradley, 2688).

Corwin (2009) analyzed the pitch contours of a group of genderqueer participants to investigate how pitch can be used to construct nonbinary identities (Corwin, 1). Corwin (2009) had 15 genderqueer participants engage in interviews as well as natural conversation in social events while being recorded (Corwin, 4). Corwin (2009) analyzed the recordings for multiple factors, including style and embodied performance–but most relevant to this study is their analysis of the participants' use of pitch (Corwin, 6). Two participants were eliminated from phonetic analysis due to poor audio quality, but from the remaining 13 participants, 9 employed what Corwin (2009) refers to as "a unique phonetic pattern that combines stereotypically feminine and masculine phonetic patterns" (Corwin, 5). Corwin (2009) states that having a small pitch range is commonly a masculine feature, while frequent use of high rise terminals (uptalk) is commonly a feminine feature (Corwin, 6). Corwin (2009) found that 9 out of 13 genderqueer participants combined these two features, using both a smaller range of pitches and high rise terminals in their

speech, which made them "sound distinctly gender ambiguous" (Corwin, 6). This study aligns with the the findings from Shar (2018) and Schmid & Bradley (2019), and shows that indexing nonbinary gender with intonation applies to uptalk specifically as well.

The demographics of the participants in my study will be discussed in detail in section 3.1, but it is relevant to the discussion of uptalk and gender to note that while all of my participants identify as women, 3 out of 25 also identify as nonbinary and one expressed uncertainty about their gender identity. From the sources in the above section, there is evidence to suggest that nonbinary gender identity–as well as gender expression for individuals of any gender–can influence how a person uses intonation in their speech. This present study will not be looking at gender identity as a variable, but rather pooling together F0 measurements of groups of participants to make generalizations about women's usage of uptalk. But it is important to keep in mind that the specifics of one's gender identity can impact how they employ intonation, and generalizations about women's speech cannot begin to encompass the intricacies of indexing gender with language. Quoting from Shar (2018): "our consciousness of what is masculine and feminine is based on stereotypes of an idealized form. These archetypes are the stagnant, pre-discursive identities that society has made normative, although most people do not meet the standards of such normative identities" (Shar, 13). In line with Shar's (2018) standards, when referring to aspects of "women's speech" in this study I am referring to how speech is used to perform femininity and not suggesting that my participants conform to a uniform gender category.

2.2.2 Uptalk and Race

Continuing the discussion of the demographics of my participants, the racial and ethnic identities of my participants can be seen in table 1 below.

(1)

Racial/Ethnic Identity	Number of Participants
Asian and White	1
Black/African American	1
Chinese American	2
Indian and White	3
Korean American	1
Latinx and White	1
Mexican American	1
Pakistani American	1
South Indian	1
White	10
White and Ashkenazi Jewish	3

While the participants represent a variety of racial and ethnic identities, the majority (52%) of my participants are white. This means that results from this study will not necessarily be applicable to all women's speech, as the data comes from majority white women's speech. In order to supplement my data and provide additional information and context that cannot be accounted for by this particular group of participants, I will explore previous research about intonation, uptalk, and race.

In addition to discussing gender in their study, Linneman (2013) discusses the race of uptalk users on *Jeopardy!* as well. Using the same episodes 100 episodes from 2009 and 2010 as well as the 49 special episodes, Linneman (2013) coded each contestant's race as white, Black, or "other," due to limited racial diversity on *Jeopardy!* (Linneman, 89). In order to reduce subjectivity in the coding process, Linneman (2013) hired nine coders and compared their judgements for each contestant (Linneman, 89). Using these classifications, Linneman (2013) then calculated rates of uptalk usage frequency as it relates to the contestants' race. Linneman (2013) found that white women were much more likely to use uptalk than Black women, with white women having a 48% change of using uptalk and Black women having a 13% change of using uptalk (Linneman, 95). Linneman (2013) hedges these results, though, by stating that there were only 6 Black women featured in the episodes (Linneman, 95).

Though there is a lack of substantial research on Women of Color's use of uptalk, there is more available information on intonational features of Black English. Tarone (1973) provides a foundation of information about how

Black English commonly has a wider range of intonation than white speech (Tarone, 29). Tarone (1973) recorded, transcribed, and analyzed natural speech from members of three communities of practice: seven Black adolescents in Chicago (mixed gender group), eight white adolescents in Chicago (mixed gender group), and the director of a Black community center (male) (Tarone, 31-32). These two groups and one individual were describes as using three different varieties of English respectively: Black English, White English, and Formal Black English (Tarone, 32). After careful phonetic and statistical analysis of each participant's speech, Tarone (1973) found that speakers of Black English used a wider range of intonation than speakers of White English and Formal Black English (Tarone, 32). Additionally, the speakers of Black English had more "level and rising final pitch contours," while Formal Black English and White English speakers used more "falling final contours" (Tarone, 33). These findings about intonation in Black English would later be named "AAVE Intonation" by Rahman (2008). Rahman (2008) notes that researchers broadly agree that both male and female speakers of AAVE (African American Vernacular English) employ more varied intonational contours than male and female white speakers (Rahman, 147). They also expanded upon this finding, showing that voices judged by listeners to sound like white women have a significantly smaller F0 range than voices judged to sound like Black women (Rahman, 148).

These findings are particularly pertinent to the present study, as these results suggest that speakers of Black English may be more likely to use uptalk in casual speech than in formal speech. Tarone (1973) also provides discourse functions for uptalk in Black English, stating that uptalk implies "that the speaker considers his utterance to be incomplete by itself and in need of some kind of supplementation. This supplementation may be a comment from the hearer or an inference he is to draw" (Tarone, 33). These suggested discourse functions of uptalk in Black English are consistent with those suggested by other researchers in section 2.1, particularly the confirmation request. Tarone's (1973) findings that speakers of Black English may be more likely to use uptalk in casual speech and less likely to use it in formal speech could be seen as consistent with Linneman's (2013) findings, as *Jeopardy!* is a more formal speaking environment and therefore researchers found less instances of Black contestants using uptalk.

Returning to the implications of the participant group for this present study being 52% white, Fought (2013) offers insight on the importance of studying race, ethnicity, and language change. Fought (2013) notes that the majority of sociolinguistic research, particularly about sound change, has focused on European-American speakers in large cities (Fought, 456). This means that many minority ethnic communities in the United States go under-researched, and therefore the small amount of research on the language of these communities can be grossly over-generalized far beyond what can actually be supported by research (Fought, 457). In the present study, with only 25 total participants, a majority of which are white, it is especially important to take into account Fought's (2013) warning to not over generalize findings and assume that trends from one community apply to all (Fought, 458). It is also important to acknowledge that the majority of uptalk research is conducted on white people and white women specifically, and it is crucial that substantial phonetic research on Women of Color is explored as well. Quoting from Fought (2013): "It is particularly important not to discourage younger scholars from pursuing the many unexplored areas of variation and change in minority ethnic communities by suggesting that there is nothing of interest to find. With luck, future research on the many communities that have not yet been studied will resolve some of the questions raised here" (Fought, 2013). Given the limited scope of the present study, my results will leave much information about women's speech uncovered and hopefully raise questions for further, more comprehensive research in on uptalk and Women of Color in the future.

2.3 Experimental Methods in Phonetics

There is ample previous research on the best methods to use in phonetics experimentation that I referred to for this project. Taboada's (2006) paper titled *Spontaneous and Non-Spontaneous Turn-Taking* offers a model of language tasks that are useful for the present study. Taboada's paper focuses on methods of turn-taking in Spanish conversations. Their research focuses on lexical signifiers of turn-taking, which is not relevant to my research, but their methods of eliciting turn-taking conversation in an experimental setting and their background information on turn-taking and floor holding are useful. Taboada is careful to distinguish between a turn and a backchannel—a turn shows intention to take the floor while a backchannel does not (Taboada, 331). This is relevant to my research because I also intend to distinguish between turns and backchannels, and my hypothesis is that uptalk may be more salient in instances where turn taking is more common and urgent than in instances where a listener is exclusively offering backchanneling.

Taboada (2006) also includes an experimental task for turn-taking that is useful for developing my methods plan. In order to elicit turn-taking in conversation, researchers used a "purposeful, staged, goal-oriented activity" (Taboada, 338). Two speakers were given a schedule with complicated commitments, and they had to verbally plan a time to "meet" that worked with both schedules (Taboada, 338). I used this concept of a goal-oriented activity to elicit back and forth speech in my experiment as well. Creating a complicated and collaborative communicative task will elicit a need to hold to floor and confirm that one's listener understands what they are saying, which will in turn elicit uptalk. Theoretically, the more spirited this conversation is the more extreme the pitch contours will become. Taboada's example of a turn-taking conversation task was used to create an even more complicated and spirited task in my own study to elicit uptalk.

Finally, the chapter by Ito & Speer (2006) titled *Using Interactive Tasks to Elicit Natural Dialogue* from the book *Methods in Empirical Prosody Research* offers insight on how to design speech tasks to elicit natural speech for the purposes of prosody research. In the chapter, the authors offer examples of speech tasks for different types of studies, many of which are not relevant to this study as they offer information on how to elicit specific words or sounds from speakers. But, one important insight they offer is that natural speech allows the appearance of more prosodic elements than scripted speech (Ito & Speer, 254). For this reason, I did not have my participants read from scripts or act out scenes as I have seen in some similar studies. Ito & Speer (2006) also give examples of speech tasks which involved physical items, which helps speakers remain on-task and can be manipulated to produce a specific situation or type of speech. Their recommendation is a "tree decorating task" in which a director gives instructions to an interlocutor as to how to decorate a tree (Ito & Speer, 245). This specific task would not work for my study, as I want the two paired participants to be equals in conversation so both of their speech can be analyzed for uptalk. However, I did use physical items to encourage active participation in the task and elicit more spirited speech.

3 Methods

3.1 Participants

The participants in this study are all women-identifying students at the Claremont Colleges (the 5Cs). Any 5C student who identifies as a woman and is a native speaker of English was eligible for this study, regardless of race, ethnicity, other gender identities, or proficiency in languages other than English. All participants were between the ages of 18-22 years.

The participants were recruited using a Google survey which was posted to multiple 5C community Facebook pages. The form asked for interested parties to give their availability and contact information, and I used these responses to create a schedule for running trials. I was then able to contact the participants and schedule them for their speech task.

Participants were divided into three experimental groups depending on the availability of their schedules, while trying to keep the number of participants in each group relatively equal. The L group contained 9 participants, the M group contained 8 participants, and the H group contained 8 participants. Details about these groups will be given in section 3.2.

At the end of their trials, each participant was given an optional exit survey. They were informed that their answers would be associated only with their participant number, not their name. In the survey, they were asked: *Do you have other identities within your gender identity that you are comfortable sharing? What is your racial and ethnic identity?* In response to the first question, 13 wrote that they identify as cisgender, 3 wrote that they identify as non-binary, one wrote that they were "not sure," and 8 chose not to give more information about their gender. All 25 participants chose to include information about their racial and ethnic identities, and their responses can be seen in table 2 below.

(2)

Racial/Ethnic Identity	Number of Participants
Asian and White	1
Black/African American	1
Chinese American	2
Indian and White	3
Korean American	1
Latinx and White	1
Mexican American	1
Pakistani American	1
South Indian	1
White	10
White and Ashkenazi Jewish	3

All 25 participants were informed of the basic tasks they would perform and the fact that there were no emotional risks to the study. They were also given a consent form to review and sign before beginning. All participants were compensated with $10 paid via Venmo and spent one hour or less on their task. All participants completed their speech tasks in a private study room.

3.2 Experimental Groups and Procedure

There were three experimental groups: high urgency, medium urgency, and low urgency (referred to as H, M, and L going forward). The independent variable in this study is the level of urgency, and the dependent variable is the degree of pitch rise. The L group contained 9 participants, the M group contained 8 participants, and the H group contained 8 participants, meaning there were 25 participants total. These numbers were chosen because 25 is a large enough group to obtain some general findings, but small enough that it was be attainable to recruit enough participants in the given time. The H group was divided into 4 pairs. Each pair completed a collaborative speech task together, which elicited lively speech and communication between the two. The participants were each given a piece of paper containing the fictional scenario below (adapted from icebreakerideas.com):

You understand the importance of team work in your job. You share ideas and responsibilities with your team members on a daily basis. In your weekly team meeting with your supervisor, one of your co-workers, Maya, takes credit for a time and money saving change in operating procedures that you know your co-worker George devised. Your supervisor, Clara, erroneously thinks Maya came up with the change and neither Maya nor George corrects the misinterpretation. The boss not only commends Maya but offers her a bonus.
Some facts:

- Maya is often talking over other people at work
- George is really quiet at work, and could use a confidence boost
- You saw Maya apologize to George privately
- Clara has asked you before not to involve yourself in others' work conflicts
- Maya has been performing great at work in general
- Clara always encourages honest feedback about her leadership
- George hates being involved in conflict; it makes him very uncomfortable
- Performance reviews are coming up, and they may be making cuts at the company
- Your team always shares ideas and discourages the concept of ownership of thoughts
- Maya has been struggling financially recently, and this bonus may help her
- George had to quit his last job because he was not respected by his coworkers
- You asked George about the situation, and he told you not to worry about it

The participants were given as much time as they needed to read over the information. Then, I randomly assigned each of the two members of the pair a position to take: either they will keep quiet and not intervene, or they will go to Clara and explain that the idea was George's. I then gave them 6 minutes to complete the following assignment: **Imagine you are both part of the consciousness of this one person in the situation, and have a conversation with one another about why your plan of action is best given the information above. Try to convince your partner that your plan is better than theirs. Both partners should argue for their side**

using evidence and explanations of their thought processes. This should be a back-and forth exchange of ideas for the duration of the 6 minutes. The goal of this task was to elicit turn-taking and the need for floor-holding between the two participants and create a sense of urgency for the speakers to be able to say their piece. The speakers were recorded using an iPhone 10 which sat on the table between them.

The M group was not divided into pairs; instead, each participant talked directly to me. I gave the participants the same fictional scenario and information. I also randomly assigned them a side to take. The M group participants' speech task lasted 3 minutes, and they were given the following assignment: **Imagine you are the person in this situation and explain to me why your plan of action is best given the information above. Try to convince me that your plan is best. You should argue for your side using evidence and explanations of your thought processes. I will be here actively listening to you, but I won't interrupt or offer my own thoughts for the duration of the three minutes.** I listened, nodded, and backchanneled to the participants, but I made it clear to them that I would not be interjecting or adding my own commentary. The purpose of this task was to see if participants will look to me for confirmation requests, but not necessarily feel to need to maintain the floor or establish turn-taking. Once again, the conversation was recorded with an iPhone 10.

The L group also be completed their task individually, but I was not be in the room for their speech task. They were given the exact same information, time to review, and instructions as the M group. The only difference was that they explained their thoughts directly to the recording device with no one else in the room. Their assignment was: **Imagine you are the person in this situation and explain to a potential listener why your plan of action is best given the information above. Try to convince a listener that your plan is best. You should argue for your side using evidence and explanations of your thought processes. You will be speaking to the audio recording only, and I will leave the room for the duration of the 3 minutes.** The purpose of this task is to see if there is no interactive element to the speech at all, whether the speakers will need to use uptalk for any of its discourse functions. The L group was also recorded using an iPhone 10, and I left the room once the recording began and returned once the 3 minutes was over.

3.3 Measurements

Once I recorded speech samples for all participants, I converted my files to .wav files and imported them into Praat in order to take measurements. Each speaker was assigned a number, so their identities remained anonymous during the data analysis portion. One participant was eliminated from the M group due to poor audio quality, so only 7 participants were analyzed for the M group.

The first step of taking measurements was to identify each instance of uptalk, as I only measured the pitch of phrases that were pre-identified as containing uptalk. I did this based on my own judgements as a linguistics student and native speaker of English, which is similar to the method of classing instances of uptalk used in Ritchart & Arvaniti (2014). I identified the instances of uptalk using my judgement of the audio in combination with the auto generated pitch contours in Praat, looking and listening for upturned contours at the end of declarative sentences.

The next step of taking measurements was to segment the pitch rise in Praat and save the labeled segment to a textgrid. It is important to segment each instance of uptalk in a uniform manner so they can all be compared to one another, so I followed the segmenting guidelines below for all of the files:

1. Start of rise is determined as the spot after the onset of the last stressed vowel (LSV) (Ritchart & Arvaniti, 332) where there is a visible rise and the pitch rises by at least 5 Hz in the following .05 seconds
2. End of rise is determined as the end of a nasal, liquid, glide, or voiced fricative (as these consonants carry pitch) which follows the LSV, or just the end of the LSV if followed by a different consonant
3. If the rise spans over multiple syllables, start with the last stressed vowel and include all following sounds until the end of the word, and use the same guidelines in (2) for the end of the segment
4. If the rise spans over multiple words, pick the word or small word cluster (4 syllables or less) where the rise begins OR the one which is at the end of the phrase (whichever bears the majority of the rise) and measure just that/those word(s) using the guidelines in (1) and (2)
5. The scaling will be the pitch in Hz at the highest point in the rise minus the pitch in Hz at the beginning of

the rise (Ritchart & Arvaniti, 332)

These segmenting guidelines were developed using previous methods from Ritchart & Arvaniti (2014), with some changes made to fit the specific needs of this study and the features of my audio files.

The pitch tracking feature in Praat is reliable, but if the audio quality becomes quiet, the speaker uses creaky voice, or there is interference from another audio source, the tracker can glitch easily. Because of this, I had to eliminate some instances of uptalk from many of the speakers because Praat could not calculate pitch for some or all of the rise. The number of tokens eliminated along with the total instances of uptalk per participant can be seen in table 3 below. Note that the participant number begins with the letter corresponding to the group they were in (L, M, or H) and that participant M5 was eliminated due to poor audio quality. The participants in H group who worked in pairs are shown with their uptalk instances grouped together. The column indicating total instances of uptalk includes the number of instances that were eliminated.

(3)

Participant	Total Instances of Uptalk	Number of Instances Eliminated
L1	17	3
L2	5	2
L3	28	9
L4	7	1
L5	18	3
L6	35	11
L7	13	0
L8	7	4
L9	14	3
M1	10	4
M2	8	0
M3	6	0
M4	13	1
M6	8	4
M7	17	2
M8	31	13
H2 (pair)	19	9
H3 (pair)	21	10
H4 (pair)	40	21
H5 (pair)	46	6

In addition to deciding to eliminate some instances of uptalk due to poor pitch tracking in Praat, I also had to make decisions about how exactly to classify uptalk as I was segmenting. Uptalk is frequently defined as a high rising terminal at the end of a declarative utterance, but there are multiple different types of high rises that can carry different meanings and discourse functions. Two types of high rises that are transparently uptalk are L*H-H% and H*L-H% rises, which are low to high rises and simple fall rises respectively (Fletcher et al. 229). An example of an L*H-H% rise can be seen in figure 2.

The rise in figure 2 above is a transparent example of an uptalk rise, as it terminates in a high rise and sounds like a question. Rises like these were consistently included and segmented during the measurement process. In contrast, there were many instances of participants using L*H-L% rises in their speech throughout their trials. These rises, while they do sound like high rises, end in a low boundary tone and therefore differ from the classic question intonation that is characteristic of uptalk. An example of one such L*H-L% can be seen in figure 3.

As we can see by following the blue pitch tracker in figure 3, after the pitch reaches its peak it begins to fall again before the end of the word. This is not an example of uptalk, because it sounds more "sing-songy" than questioning and doesn't terminate in a rise. Therefore, I did not segment these rises during measurement.

Once all of the uptalk instances were segmented and saved into textgrids, I went through each file and checked my work for a second time to ensure that each segment was done correctly. Next, I ran a script called ProsodyPro (Xu, 2013) through all of the textgrid files to extract the maximum F0 value from each segment. This measurement

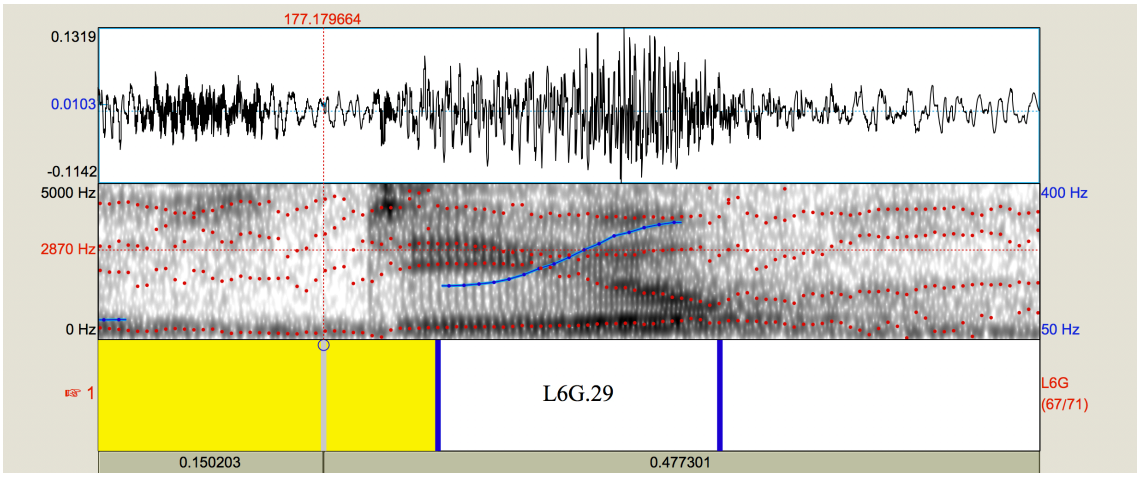

Figure 2: A segmented rise on the word “due”

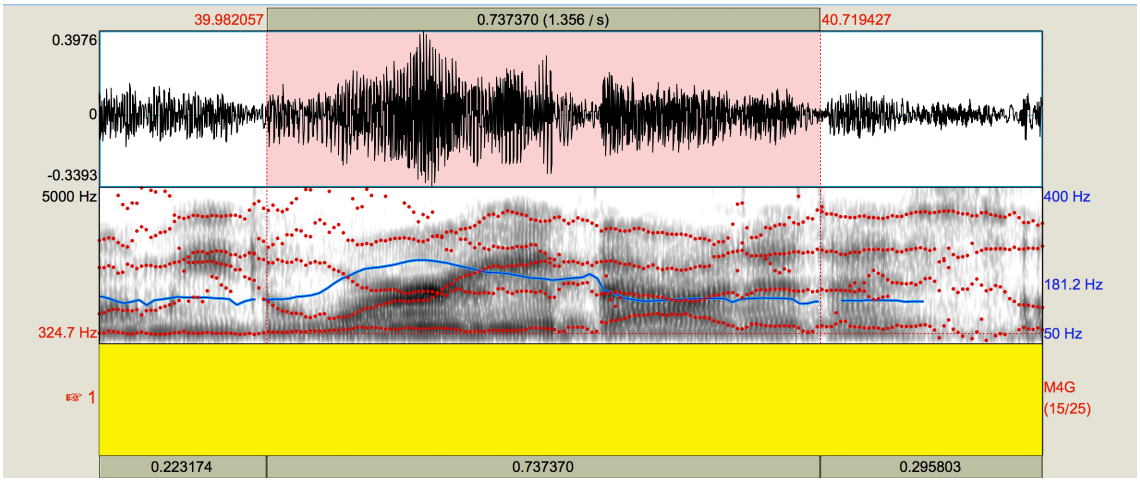

Figure 3: A rise-fall on the phrase “worry about it”

represents the highest fundamental frequency that the speaker produced during the segmented section, commonly known as the highest pitch. In order to calculate scaling, I also needed the F0 value at the start of the pitch rise, which was also the start of the segmented section. ProsodyPro does not extract this information, so I found these values manually. I did this by going into each file and clicking on the leftmost boundary of the segment and using the "get pitch" feature in Praat. This produced the F0 value at the leftmost boundary, which is the pitch at the beginning of the rise. After I extracted all the initial pitch values, I imported them into a spreadsheet along with the maximum pitch values for each instance of uptalk.

Once the values have been extracted and organized, I used them to calculate the scaling for each instance of uptalk. The scaling is calculating by subtracting the pitch in Hz at the beginning of the rise from the pitch in Hz at the highest point in the rise. This was an easy calculation that was done in Google Sheets. I then organized these values by experimental group, giving me an array of scaling measurements for each of the three groups. At this point, the data was ready for statistical analysis.

4 Analysis and Results

Using SSPS, I conducted a one-way analysis of variance (ANOVA) test on the scaling values between the L, M, and H groups, as well as post-hoc comparison tests using Tukey HSD to determine statistical significance between groups. SPSS also produced descriptives for each experimental group. First, we will look at the descriptives in tables 4 and 5. Note that the average number of tokens column in table 4 is calculated using the number of tokens after some were eliminated due to poor audio, while the number of instances of uptalk in table 5 includes all instances.

(4)

	Num. Tokens	**Mean Scaling (Hz)**	**Standard Deviation (Hz)**
L Group	108	62.33	38.53
M Group	68	50.13	34.23
H Group	80	32.72	24.54
Total	256	51.94	36.53

(5)

	Num. Participants	**Total Instances of Uptalk**	**Avg. Instances Per Participant**
L Group	9	144	16
M Group	7	93	13.29
H Group	8	126	15.75
Total	24	363	15.01

As we can see from table 4, there does appear to be a difference in the mean scaling between the three groups. Notably, the scaling is highest for the L group and lowest for the H group, which is the opposite of the hypothesized effect. However, the M group had a mean scaling which fell in between the L and H groups, which is in line with the original hypothesis and experimental design. Also of note is the large standard deviations, which show that the pitch rises varied within groups. It also appears that all three groups uptalked relatively equally frequently, with the L group having a slightly higher average per person and the M group having a slightly lower average per person. To determine if the difference between the mean scaling is significant despite the deviations, we can examine the results of the one-way ANOVA test. The results of this test can be seen in table 6 below.

(6)

	Sum of Squares	**df**	**Mean Square**	**F**	**Significance (p-value)**
Between Groups	55365.324	2	27682.662	24.583	$1.7529E^{-10}$

The key value in this table is the significance or the p-value. The p-value for the test of variance between the groups is far less than 0.05, meaning the variance is significant. This means that despite the large deviations, there is a significant difference in the scaling between the groups–but this test does not tell us *which* groups are significantly different from each other. To determine that, we need to perform post-hoc Tukey HSD tests between each of the groups to see which groups are significantly different from one another. The results of these tests can be seen in table 7 below.

(7)

(I) Group	(J) Group	Mean Difference (I-J) (Hz)	Std. Error	Significance (p-value)
L Group	M Group	17.20	5.19	0.003
L Group	H Group	34.61	4.95	$5.1691E^{-9}$
M Group	H Group	17.41	5.54	0.005

Table 7 shows that each of the three groups is significantly different from each other in terms of scaling. The post-hoc tests perform simple variance tests between each group, which can be seen in the (I) Group and (J) group columns in the chart. The most pertinent measurement from these results again are the p-values, which show if there is a significant difference for each group comparison. All p-values are lower than 0.05, meaning there is a significant difference in the scaling between all three L, M, and H groups. The final important test to address is the test of homogeneity of variances. This test measures if the variance that exists in each group is significantly different between each group. The results of this test can be see in table 8 below.

(8)

	Significance (p-value)
Based on Mean	0.000047
Based on Median	0.000082
Based on Median and Adjusted df	0.000084
Based on Trimmed Mean	0.000043

As we can see from table 8, the results of all the tests of homogeneity of variances are significant as their p-values are less than 0.05. This indicates that the groups do not have the same level of variances–some groups have more variance within them than others. This makes logical sense due to the nature of the scaling measurements–the measurement is bounded on the left as there has to be a rise of at least 5 Hz for the instance to count as uptalk. But, there the measurement is not as strongly bounded on the right, as a rise could go up by any amount of Hz. This creates an effect where those groups with lower means would have less variance, and those with higher means are less restricted and therefore have higher variances. In order to account for this difference in variance, I ran an additional test using non-parametric techniques which confirmed the results of the initial ANOVA test. The results of the non-parametric test reflected the original pattern of results, strengthening the significance of the difference in scaling between the L, M, and H groups.

To summarize the results of the statistical analysis, I found a significant difference between the scaling values for the groups groups, and all three groups were statistically significantly different from each other. Contrary to my initial expectation, the L group had the highest mean scaling value, and the H group had the lowest mean scaling value. See the graph in figure 4 for a visual representation of the mean scaling values for each group.

5 Discussion

From the descriptives and the results of the statistical tests, we see that the L group employed more extreme pitch contours (and therefore had the highest mean scaling), the H group employed the least extreme pitch contours, and the M group landed in the middle. This is the opposite result of my initial hypothesis–I thought that the H group would have the highest mean scaling because they completed their speech task in pairs. Because uptalk is a social communicative tool that is used for floor holding and confirmation requests, I hypothesized that back and forth conversation would elicit higher scaling numbers than speech that took place alone in a room. But the results of this study show the opposite.

Guy & Horvath (1986) provide information about the uses of uptalk that could contextualize these results. In a study they conducted about the distribution of uptalk use, they found that uptalk is used during lengthy, complicated narratives to maintain attention (Guy & Horvath, 43). Though the speech task in this study was more of an explanation than a narrative, the participants in groups L and M did have to speak for a long period of time (3 minutes) without interruption. Perhaps their use of extreme pitch contours reflected that they were more aware of the amount of time they were speaking the subsequent need to maintain their listener's attention than the H group, who were taking turns and therefore speaking for shorter amounts of time at once. Additionally, the L group was alone in the room and had no one to respond to their confirmation requests. Perhaps their knowledge that someone

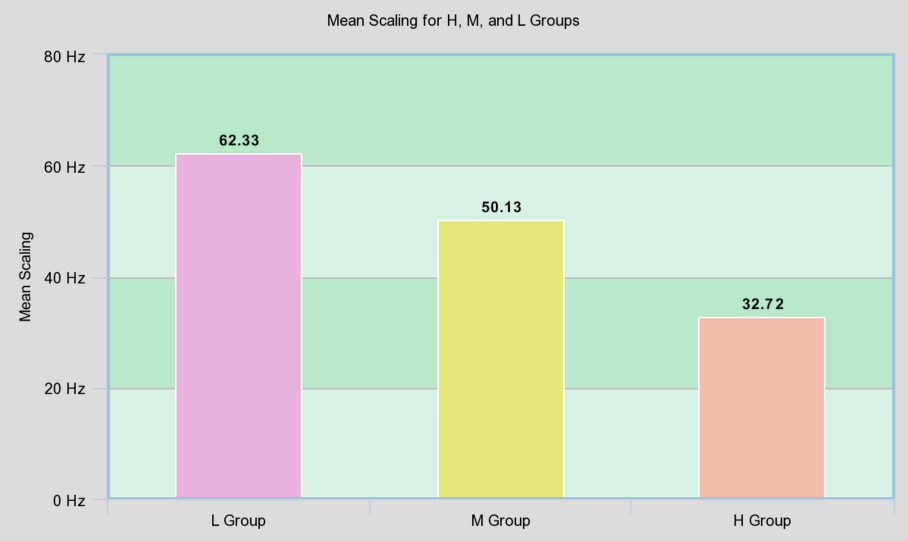

Figure 4: Bar graph of scaling means for H, M, and L groups

would listen to their speech later combined with the lack of input from a listener prompted more extreme uptalk pitch contours and they subconsciously tried to maintain listener attention. Though Guy & Horvath's (1986) information helps contextualize these results, their study was on the frequency of uptalk use and not on the scaling of uptalk.

Research on uptalk scaling is currently sparse, but we can turn to Ritchart & Arvaniti (2014), the study which laid the groundwork for much of this present project, for some more insight into how uptalk scaling changes based on discourse function. Ritchart & Arvaniti (2014) found that uptalked phrases that served confirmation requests bore the most extreme scaling as compared to floor holding, question, and other unspecified statements (Ritchart & Arvaniti, 333). Though floor holding statements still had considerably large scaling means, the means were smaller than those for the confirmation request statements, meaning that the need for floor holding does not produce as extreme pitch contours as the need for confirmation requests. This could potentially help explain the results of this present study, as one could argue that the main discourse function of uptalk in the H group was floor holding as the pairs were engaging in back and forth conversation and were in disagreement with one another. On the other hand, the L group did not have any need to hold the floor (as they were speaking alone in a room), maybe perhaps still felt the need to use uptalk for the purpose of confirmation requests as they subconsciously knew that someone would be listening to their recording later. With no one to offer backchanneling in response to their requests, (as the M group had), maybe this encouraged the L group to employ even more extreme pitch contours to compensate for the lack of confirmation. Perhaps the H group did not have this same effect, as they were mainly employing uptalk for the purpose of floor holding which has been shown to create smaller scaling means (Ritchart & Arvaniti, 333). It is also possible that the conversations in the H group did not feel very high-pressure to the participants–maybe they treated it more as a collaborative task rather than a debate. This could affect their motivation to use uptalk or the degree to which they employed it.

6 Conclusion

The theories presented in section 5 are simply conjecture and cannot be proven by this study, and further research would be necessary to fully discover the explanation behind the results of this project. But this study has expanded upon the previous research on uptalk and particularly on uptalk scaling. This study built on the foundation laid by Ritchart & Arvaniti (2014), who showed that uptalk scaling changes depending on discourse function. My study examined whether uptalk scaling also changes depending on conversational context, and the results have suggested that it certainly does. This study contributes to the existing research on uptalk, pitch scaling, and women's speech, and created new questions about conversational context and uptalk that could be addressed by further re-

search.

6.1 Further Research

Further research on this subject could look into the types of rises that I excluded from measurement, namely L*H-L% rises. These rises were used frequently by my participants, but since they ended in a falling intonation, I did not count them as uptalk for this study. Further research could investigate the communicative functions of these rises and how they differ in use from uptalk rises.

There is a significant deficit in prosody research on Women of Color, and as this study had a majority white participant pool it was not able to contribute to that field. Further research could investigate the findings of this study specifically on a participant pool of Black women, as there is currently interesting and conflicting information about Black women's use of intonation and uptalk. Tarone (1973) found that Black men and women employ broader pitch ranges and more high rising terminals than white speakers, but Linneman (2013) found that Black women use uptalk less than white women. These findings could be better contextualized and explained through a study similar to this present study but with a focus on race.

Additionally, further research on uptalk and gender which goes beyond the gender binary is important to fully understand how intonation is used to index gender as well as how it is perceived by others. A study on the discourse functions of uptalk as used specifically by nonbinary individuals could expand upon the present study as well as the findings of Corwin (2009), Schmid & Bradley (2019), and Shar (2018).

Another interesting expansion of this study would be to code what happens in the discourse after the use of uptalk. How successful is uptalk for eliciting confirmation, backchanneling, holding the floor, and preventing interruption? A study which investigated these questions could provide more information about how uptalk in used in conversation and the way it fits into the discourse context surrounding it.

Finally, further research could investigate the reasons behind the results of this present study. Perhaps a study that focused on how and why uptalk is used by speakers when they are alone could illuminate why the mean scaling for the L group was the largest of the three groups. Research on uptalk used by a speaker when they are not going to be interrupted or take turns, such as in a presentation context, could also be an interesting addition to these results.

7 Acknowledgements

I would like the thank Professor Jeffrey Geiger, Professor Mike Diercks, Professor Robin Melnick, Sara Goring, and my entire Senior Seminar class for advising me and providing me with feedback throughout this project.

8 References

Banuazizi, A., & Creswell, C. (1999). *Is that a real question?: Final rises, final falls and discourse function in yes-no question intonation.* www.sfu.ca. Retrieved September 28, 2021, from http://www.sfu.ca/ hedberg/Banuazizi_questions.pdf%20.pdf.

Boersma, Paul & Weenink, David (2021). *Praat: doing phonetics by computer* [Computer program]. Version 6.2.01, retrieved 1 November 2021 from http://www.praat.org/

Chambers, J. K., Schilling-Estes, N., & Fought, C. (2013). Ethnicity. In *The handbook of language variation and change.* essay, John Wiley Sons.

Corwin, A. I. (2009). Language and gender variance: Constructing gender beyond the male/female binary. *Electronic Journal of Human Sexuality*, 12, 1–11.

Fletcher, J., Stirling, L., Mushin, I., & Wales, R. (2002). Intonational rises and dialog acts in the Australian English map task. *Language and Speech*, 45(3), 229–253. https://doi.org/10.1177/00238309020450030201

Guy, G., Horvath, B., Vonwiller, J., Daisley, E., & Rogers, I. (1986). An intonational change in progress in Australian English. *Language in Society*, 15(1), 23–51. https://doi.org/10.1017/s0047404500011635

Tyler, Joseph C. "Expanding and Mapping the Indexical Field: Rising Pitch, the Uptalk Stereotype, and Perceptual Variation." *Journal of English Linguistics*, vol. 43, no. 4, 2015, pp. 284–310., https://doi.org/10.1177/0075424215607061.

Linneman, T. J. (2013). Gender in *Jeopardy! Gender & Society*, 27(1), 82–105. https://doi.org/10.1177/0891243212464905

Rahman, J. (2008). Middle-class African Americans: Reactions and attitudes toward African American English. *American Speech*, 83(2), 141–176. https://doi.org/10.1215/00031283-2008-009

Ritchart, A., & Arvaniti, A. (2014). The form and use of Uptalk in Southern Californian English. *Speech Prosody* 2014, 7. https://doi.org/10.21437/speechprosody.2014-54

Schmid, M., & Bradley, E. (2019). *Vocal pitch and intonation characteristics of those who are gender non-binary.* International Phonetic Association. Retrieved November 19, 2021, from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2019/papers/ICPhS_2734.pdf.

Shar, E. (2018). *Expressing gender and queer identity with intonation: Variation amongst masculine, neutral, and feminine aligning Afab Queer Speakers in DC* (dissertation).

Sudhoff, S., Ito, K., & Speer, S. R. (2006). Using Interactive Tasks to Elicit Natural Dialogue. In *Methods in Empirical Prosody Research* (pp. 229–257). essay, Walter de Gruyter.

Taboada, M. (2006). Spontaneous and non-spontaneous turn-taking. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 16(2-3), 329–360. https://doi.org/10.1075/prag.16.2-3.04tab

Tarone, E. E. (1973). Aspects of intonation in Black English. *American Speech*, 48(1/2), 29. https://doi.org/10.2307/3087890

Xu, Y. (2013). ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), Aix-en-Provence, France. 7-10.