1 Introduction

My friend, Cathy Barkett once related that, based solely on a phone conversation, she thought that a woman with whom she was speaking was "heavy-set." Although she could be talking about the voice, it became evident that she was instead talking about the woman's body. Nevertheless, her perception was based solely on the woman's voice. She felt justified in her intuition when her husband, who knew the woman in question, verified her assertion that the woman was indeed "heavy-set." Because these sorts of assertions and verifications are commonplace in daily life, I believe they are one source of stereotypes. That is, stereotypes are built from positive feedback (like affirmations from others) on generalizations derived from apparent correlations between observed (speech) features and observations of the people who have them. This, at least, must be a way people build unconscious stereotypes based on voice alone, but naturally, these stereotypes are not limited to judgments of weight and/or body type. They exist based on voice alone for race, gender, class, education, and sexual orientation. It is the final of these that I will be studying in great detail in the following pages.

points to areas in the study of gay and lesbian language use where more work needs to be done:

> In addition to more research that measures pitch (average, range, and variability) with more subjects, these questions for future studies should include: (1) What exactly are the phonological features that differentiate gay and heterosexual men, and in what situations might these features be more prominent?… (p. 53)[1]

---

[1] Citations in the form (p. X) refer to another work while citations in the form (see page X) are cross-references within this document.

In response to this call, I created a study on this topic with more subjects in my experiment (seventeen speakers and thirty-two listeners) than have been previously seen in studies on this topic. And although I think there is a faulty premise in Jacobs' request, this study does deal with phonological (and phonetic) features and the different situations in which they occur. The fault I find with Jacobs' request is in the notion of "the phonological features that differentiate gay and heterosexual men." To be blunt, there are no phonological features that *differentiate* gay and heterosexual men.[2] Sexual orientation does not determine or even imply certain phonological features as this request assumes. My idea runs counter to the assumption in that gay-sounding speech features really only exist for the purpose of announcing to listeners the sexual orientation of the speaker. For Zwicky, gay-sounding speech is *performed*. As I will assert later, I am highly skeptical of this claim because there are a number of gay-sounding straight-identified men and straight-sounding gay-identified men. Thus, it follows that these speech features are not (consciously) performed, because their existence or non-existence does not always accurately reflect the identified orientation of the speakers that have them. There is nothing about an individual's sexual orientation that entails that individual's having certain features.[3] In other

---

[2] I do not understand why people mix levels in the form: "gay and heterosexual." It should be "homosexual and heterosexual"— the medicalized terms— or "gay and straight"— the colloquial/general usage— not "gay and heterosexual." It seems as irrational as saying "Group A and Group 2" or "Group A and Group _."

[3] Even if listeners' evaluations of speaker's sexual orientation tend to be accurate, that fact only fosters a self-perpetuating situation wherein the stereotypes are regarded as universally true. The point here is that since there are exceptions of the type distinctive straight-sounding feature on gay male and gay-sounding feature on straight male, making the absolute one-to-one mapping

words, one's sexuality does not determine how one speaks, so Jacob's is not an appropriate research question, as stated. What is possible to study, though, and is still just as interesting, are the perceptions, the social categories constructed regarding types of speech. In this study, I will therefore be focusing on the response of listeners:

- What do they hear?
- What preconceived notions of gay-sounding and straight-sounding speech do they have?
- How and to what extent does what they hear fit into the categories gay- and straight-sounding that they carry in their minds?
- In addition, how do the responses of listeners compare to the self-evaluations of speakers regarding voice and sexual orientation identification?

These questions form the basic framework of inquiry for this study. I believe the answers are to be found, in part, in measurements of pitch range and variability.

This study builds on the work Gaudio 1994, Crist 1997, and Rogers and Smyth 2001, and is designed such that the results of this study can be compared to those of the previous studies, especially those of Gaudio. It also questions the usefulness of extensive inferential statistical analysis for studies with sample sizes that are too small.[4] Further, it questions whether the categories used in

---

false, the search for deterministic gay-features (as opposed to features that listeners agree are gay-sounding) seems impossible.

[4] The commonly used t-test, for example, requires both a sample size of 50 and a normal distribution for reliable conclusions (Woods, Fletcher & Hughes 1986, p. 106), neither of which is ensured by Gaudio and Crist, both of whom do extensive statistical analysis on their small sample sizes (eight speakers, thirteen listeners and three speakers, respectively).

these studies are really useful and/or faithfully reflect the reality of the speakers' lives, that is, this study emphasizes the difference between stating "this is a feature of 'gay' speech" and "this is a feature which *listeners*[5] *identify* as 'gay' sounding." I believe the latter is a possible avenue for inquiry while the former is not possible. This is a conceptual difference that has not received sufficient attention in the literature, since the two are usually conflated as "gay features."

Therefore, the basic hypothesis of this research is that listeners can judge speakers based on phonological and phonetic features of their voices. Though data was collected to allow analysis of many features and many possible listener judgments, the results given in this paper are largely restricted to pitch features and listener judgments of sexual orientation. My findings verify Gaudio's claim that listeners can reliably identify the sexual orientation of the speakers based solely on voice recordings, and that pitch properties in speakers' reading voices do not give strong indications to listeners whereby they can make accurate judgments of sexual orientation. The most important and interesting new finding is that gay-sounding speakers sound gayer when speaking freely than when reading, and straight-sounding speakers sound straighter when speaking freely than when reading. This is important because it suggests that features that allow listeners to judge sexual orientation are more pronounced when speaking freely than when reading, and that the features are not necessarily absent when reading, just less extreme. This explains why the results of this study show many measures where pitch properties correlate with listener judgments of free response recordings, while judgments of read passages show reduced or no correlation. Through this lens earlier literature can be reevaluated to the extent

---

5 This is within the framework of listeners' preconceived notions of what 'gay' sounds like.

that small correlations based on read passages could reflect larger correlations that would exist in freer speech registers, if these had been tested.

This study looked at a small group of speaker- and listener-subjects, all of whom were undergraduates at Harvard College. It was not designed to make broad inferences about society in general, and I was therefore reluctant to use the aforementioned inferential statistical analysis on my data[6]. (I would prefer to use statistical analysis to ensure linked variables are not due entirely to chance.)[7] Rather, I wanted to pursue a more descriptive analysis of this small group in the hopes of finding more fruitful and meaningful results. In the end, I yielded to using methods similar to those of previous studies, while making improvements like more reliable statistical analysis and other additions explained below. Given that this is one of the first studies of its kind, I chose to limit my pool of subjects to a smaller community, thereby maximizing the potential to draw conclusions relevant to this particular community. The study also serves as a starting point and model for other community-based studies or more far-reaching studies in the future. Although my subjects come from many different places in the country, they all will eventually spend four years here in Cambridge, interacting in a speech community. While this grouping is imperfect in that it ignores the differences in speech patterns that can derive from age, class, and race differences, among others, it is designed to allow for making reasonable

---

[6] This is part of a larger criticism of the sociolinguistic studies I have read: the sample sizes are usually much too small to make the sorts of claims based on statistical inference that are made. I understand that for sociolinguistic studies having the sort of resources and time for large sample sizes is often a practical impossibility. But if that is the case, they should not be making broad generalizations based on risky statistical conclusions.

[7] This is the appropriate use of statistics for sociolinguistic studies with small sample sizes.

generalizations, especially with a relatively small sample size. In the end, my primary interest lies in the judgments of the speakers, not in the speakers themselves, so much of the classification data about the actual orientation of speakers is less important in this case. I am not asking listeners to answer the question literally, “Is Speaker X gay?” but instead interpret, “Does Speaker X sound gay?” In other words, “If you have a preconceived notion of what ‘gay’ sounds like, what is your notion? Does this person fit it? If so, to what extent, and how/why?”

2 Background

is the first sociolinguistics study to systematically investigate the speech properties of gay and straight men. His main research question, “What does it mean to say that a man ‘sounds gay,’” is similar to mine, though much more in line with the stated request of Jacobs 1996 (see page). He measures pitch range and pitch variability in order to determine whether or not these are among the reasons listeners judge a given voice to be gay- or straight-sounding. In his experiment, Gaudio recorded eight male graduate students reading two passages: the first from a dry accounting text, used originally in , and the second a dramatic monologue from the play *Torch Song Trilogy* (). He also then recorded a forty-five minute interview with each, asking about personal background and views on a number of gay-related topics. To prepare the recordings for listener evaluation, the clips of the passage readings were randomized within the groups for each passage and compiled on a sample tape. Gaudio also prepared a scoring rubric for his listener subjects, which included scales (from one to seven) of straight-gay, effeminate-masculine, reserved-emotional, and affected-ordinary. Then he played the clips for a group of thirteen undergraduate volunteers to have them record their judgments. Based on his data analysis, Gaudio concludes that “overall pitch range and pitch variability do not by themselves crucially affect whether or not a man will be perceived as ‘sounding gay’” (p. 53). But he does note that “(1) listeners had fairly consistent judgments about which speech sounds ‘gay’; and (2) listeners’ judgments were largely accurate with respect to identifying male voices as belonging to gay or straight men” (pp. 53-54). He also found that correlations involving pitch

variability suggest a role in influencing listeners' judgments, "but only in concert with other phonological factors" (p. 54).

It is interesting to note that although Gaudio is sometimes cautious when using categorical definitions, for example "gay," "masculine," "feminine," and "effeminate," he finds them a useful tool in producing a manageable experiment. In addition, "features of gay speech" and "features of speech that listeners judge as gay sounding" should be kept separate, but it is sometimes hard to verify if that is being done. Gaudio finds that listeners often pick out which voices belong to gay people, but in order to make this comparison, he temporarily reverses his former caution with using such binary groupings and considers each speaker as exclusively gay or straight for both speakers' self-identification and listeners' judgments. In the original paradigm, if *1* denotes "extremely straight-sounding" and *7* denotes "extremely gay-sounding,"[8] allowing (at least in some limited sense) levels between the extremes, Gaudio's comparison effectively reinstates an absolute gay-straight dichotomy, *1-4* being lumped into straight and *4-7* being lumped into gay. From his earlier exposition of issues of gay male identity, it is obvious that Gaudio does not believe solely in this oversimplification as a reflection of reality. I agree that this comparison is only of limited use because the speaker's actual orientation and identification is not necessarily connected to a listener's judgment. (If it were, it would not allow for the listener-judged gay-sounding straight men and straight-sounding gay men[9].) It seems that Gaudio's

---

[8] This scale has a long history in the study of sexuality within sociology. It is identical to the one from , p. 638.

[9] Rogers, as quoted in , speculates that perhaps less than half of gay men would be judged gay-sounding, adding, "It was interesting that the straightest-sounding voice in the study was in fact a

goal is to identify features that gay men might actively employ in order to sound more gay.[10]

My other major qualm with involves sample size and statistical analysis. Besides specifying having used Analysis of Variance (ANOVA) to determine correlation strength, Gaudio fails to specify completely how he calculated the rest of his statistical results. This may seem trivial, but it is a valid cause for concern in sociolinguistic studies since the sample sizes are often small (small in a statistical sense is under fifty). Given the small data set, the conclusions that Gaudio draws should be regarded as risky at best. In spite of these faults, however, Gaudio's work is to be commended for being, as he states, the "initial attempt toward the study of gay male speech" (p. 43) in a quantitative matter.

Following up on Gaudio's research on pitch properties, hypothesizes that the gay male stereotyped voice involves primarily the systematic lengthening of certain consonants (see page). Instead of having two passages read "naturally" by subjects *à la* Gaudio, Crist has subjects read one passage twice, once in their "normal" voice and once in a performed gay stereotype (he calls it "queeny") voice. Crist uses three subjects in each of two experiments: one gay man and two straight men in the first; three gay men for the second. He reports that speakers may or may not have /s/-lengthening, but when it occurs, /#sp/ and /#sk/ do so in tandem. (/#st/ does not participate in this lengthening pattern, and

---

gay man and the sixth gayest-sounding voice was a straight man. It's quite ordinary of gay men to not sound gay and every now and then you find a straight man who sounds gay" (p. 1).

[10] In contrast, my study is more focused on the listener response and the linguistic content of the commonly held perceptions about gay male speech (and, conversely, those of straight male speech.)

/#l/ generally follows the pattern of /#sp/ and /#sk/ but does not for specific words.)[11]

Crist also acknowledges the fact that although Gaudio did not find significant results for F_0(pitch), it may still play a role and should be studied further. I reiterate even more strongly for Crist the trepidations expressed above regarding Gaudio's small sample size. A data set of three is simply too small to do any valuable statistical analysis, even to verify the validity of the distribution of data. The second problem I have with Crist is his, frankly, bizarre experimental setup. In order to inquire into the phonetic properties of segments in "gay speech," he has his subjects perform "the queeniest, most flaming gay stereotype [they] can do." This seems akin to having various speakers of English "perform" a British English accent in order to study the phonetic differences between, say, American English and British English. British English has been studied so much that this hypothetical study *could* be interesting to find out the phonetic properties of American stereotypes of British speech. On the other hand, studies of "gay speech" or "gay-sounding speech," are extremely few. If Crist were interested in researching the properties of stereotypes of gay speech, the setup of his experiment would be perfect. But instead he infers "gay speech" features from subjects' performed stereotypes. There is a distinct incongruity between method and goal.

---

[11] The phonemic notation /#X/ signifies a string of segments (X) at the beginning of a word (#).

Finally, there is the ongoing research of Rogers and Smyth.[12] They have a broad research agenda, including looking at a number of phonetic characteristics that may be involved in listener judgments of “gay-sounding.” They claim to be investigating two common stereotypes about “gay-sounding” voices: “One is that people can recognize gay men by the way they speak and the second is that they sound effeminate” (, p.1). In a recent study, they recorded twenty-five men, seventeen of whom were gay, and had forty-seven listener subjects rate the speakers as gay or straight. The listeners were correct in identifying gay speakers sixty-two percent of the time. Two of the phonetic cues that they have identified are prolonged duration of /s/ and /z/, but the difference is very small, only a matter of milliseconds. Rogers asserts that gay men may be imitating female speech patterns[13], but he also notes that the fact that some pre-adolescent boys have gay-sounding voices calls into question the assumption of imitation of female speakers and acquisition of their speech features. He adds that women in the environment of every boy could serve as speech role models, for him a possible source of gay-sounding characteristics. He is also interested in finding out if gay men adopt (more) gay-sounding speech patterns after coming out.

---

12 My first knowledge of their work came from , a talk presented at the *New Ways of Analyzing Variation 30* conference. I received a copy of the script from Smyth, and he requested that I not cite from it. He and Rogers have been working on this topic since 1999 (and are still working on it), but have not yet published their complete findings. The information here is based on two reports of their current progress, and .

13 Though, without studying female patterns of speech, this is not a responsible claim, only serving to perpetuate folk models. This is the same spurious claim as (p. 204), which echoes .

The fact that nowhere near one hundred percent of listeners correctly identified the gay speakers is problematic for a few reasons: it implies that maybe listeners are not listening to common cues or that gay-sounding features do not always belong only to gay men. Although nothing is known about the specific statistical methods they employed, regardless, their sample size is large enough to start drawing more reliable conclusions, which is a step in the right direction. However, Rogers' repeated assertion that gay-sounding equals feminine-sounding is troubling because it is not substantiated by the data in this study. (i.e. he has done no experimental comparison of gay speech and female speech.) His claim leaves no room for gay or straight men who do not feel particularly feminine, but still sound gay (Why would these groups look to female speech role models when growing up and developing their speech patterns?).

On the other hand, reject the research of Crist 1997, Gaudio 1994 and who

> 'focused on identifying the linguistic features that constitute a style of speaking often referred to as 'gay speech' or the 'gay accent,' because of the implication that it assumes there is one gay way of speaking. This is an affront because it "erases the diversity within the 'gay community,'" a diversity of "myriad social groups such as leather daddies, clones, circuit boys and gay activists"' (p. 4).

Although there may be some truth in their claim, the studies they assail are seminal works in the area. There really was no linguistic study of this group of speakers beforehand. Therefore, treating the group as homogenous is better

than not treating the subject at all, and it gives later researchers something to work from. Later researchers can study the topic from another point of view, that of the listeners, or not regard the topic in terms of a monolithic “gay speech” or “gay accent.” In that regard, their complete rejection is not entirely valid.

Researchers now are standing on the shoulders of these giants, even if they disagree with the framing of the research, because nonetheless the data is still valid and still can be compared with later studies. When one studies “gay language,” though, one is not studying a set limited to gays alone nor simply a set including gays; one is studying is the perceptions of the interlocutors. Nothing definite can be said about the speaker himself by his voice, that is, a man is not necessarily defined by his speech patterns, nor do his speech patterns necessarily define him (this is not to discount the role that performative speech styles play in every person’s speech acts), but what one *can* study is the content of the perceptions of the voice. If speakers listen to a voice and mark it as gay-sounding, they are obviously hearing something in the voice, likely a combination of lexical, phonetic, and discourse levels among others—especially in the majority of speech interactions where the voice is not the only method of communication (e.g. paralanguage, dress, location, etc.).

So what needs to be studied, because it is what *can* be studied, is the descriptions the listeners make for the speakers and the cues that are given to create those categories. I strongly disagree with Wong, Roberts and Campbell-Kibler over the issue of diversity within the gay community for sociolinguistic studies. Although diversity no doubt exists, I do not believe that American society at large,[14] and even the majority of the members of the gay community, would

[14] This study does not attempt to study a cross-section of American society but a small subsection, that of the undergraduates of Harvard College.

necessarily make the sort of in-group distinctions that they point to based on speech features. This seems to me quite obvious for society at large, and possibly true for members of the gay community as well.

Building on the work of Crist, Gaudio, and Rogers, Smyth and Jacobs, this study principally investigates the phonological and phonetic variables in “gay speech.” After my listener subjects listen to speakers’ voices, and with only the voice to rely upon, they then report their construction of the individual. I have noted people relating stories describing the construction of a body that goes with a voice. It goes without saying that the speaker and listener are not acquainted, a situation which can easily happen on the telephone. The story of Cathy Barkett (see page 1) is a perfect example of this situation and the resultant reinforcing of stereotypes based on voice.

When a stereotype is built, it is not necessary to be in a voice-only environment, like that of the telephone. People unconsciously take note of how people speak across a wide variety of human categories (common, sometimes useful categories are labels like short/tall, thin/fat, Black/White/Hispanic/Asian, gay/straight, etc.). These help build mental databases of how a person with a given label talks, yet none of this is necessarily done through conscious decision-making. In this experiment, it is assumed that these categories exist and are commonly used in a given speech community. With new data being collected unconsciously every day in face-to-face interactions and positive feedback given after other interactions (a third-party verifying the existence of the assumed characteristic, for example), it seems these stereotypes would become quite ingrained in a member of a society. This seems like another manifestation of the concept of enculturation that anthropologists use to explain how children become members of the society in which they are raised ().

(cited in) makes a critique that is in keeping with the framework of this study: he notes that “In what [Kulick] calls ‘Darsey’s theorem’ [Kulick] condemns the overly simplistic methodology of matching sexual orientation to linguistic practice and claiming success: ‘The fact that gays do X does not make X gay’” (p. 2). That is precisely why I cannot identify specific features in listener-judged gay-sounding speech as “gay features.” I can only identify features that are common to the speech listeners identify as “gay sounding” and call them components of the shared model of a “gay speech” stereotype. One flaw in this approach, which unfortunately is not resolved here, is that if there are multiple, even discrete, “gay speech” stereotypes, the methodology here does not distinguish them and does not provide means by which they can be distinguished. Regardless, I choose to focus on listener judgments based on stereotyped categories, I believe it is a right step in one possible direction, and its focus takes into account the worries of Kulick.

3 Methodology

In formulating the speaker survey, I consulted a number of sources but relied most heavily on . I hoped to capture a number of speech levels for each speaker in my method of elicitation. I sent out advertisements over the e-mail lists of the Harvard College undergraduate houses to get subjects. Each subject made a one-hour appointment with me with the understanding that they would be participating in linguistic research. In the end I had seventeen speakers as subjects. I conducted the study in the phonetics lab of the linguistics department. Upon arrival, I presented each speaker with the first page of a packet containing passages to read and a written survey.[15]

The recordings were performed on an Apple iBook laptop computer, using the recording program Amadeus (version 3.5.1) by Martin Hairer. I also used a Sony lapel mic (model ECM-T145), which interfaces with the computer through an iMic USB adapter. I recorded at 11kHz.

3.1 Recorded data

In selecting passages for the speakers to read, I wished to be as conservative in selection as possible, which manifested itself in selecting passages that had been used in previous studies of the same nature. In my early research I found Gaudio's study to be most similar to mine, so I first decided to

---

15 I gave subjects sheets from the packet one at a time so that they could not look ahead. See page for further justification; but in short, I needed to give subjects different pages depending on their previous answers.

employ the passages he had. By doing so, I would have the option of doing the same measurements and duplicating his study to verify the results as well.

On the first page of the packet, the instructions to the speakers were as follows: “Please read over silently the following selection from a textbook. When you’re done, I’ll record your reading it aloud.” The first passage, which Gaudio also used in his study, is a dry accounting text. The text used for the first recording was as follows:

> “Fair presentation of financial affairs is the essence of accounting theory and practice. With the increasing size and complexity of American business organizations and the increasing economic role of government, the responsibility placed on accountants for presenting fairly the results of business operations is greater today than ever before. If accountants are to meet this challenge fully, they must have a logical and consistent body of accounting theory to guide them. This theoretical structure must be realistic in terms of the economic environment and designed to meet the needs of the major users of accounting information.”

This text was originally taken from in a study of F_0 in tone versus stress languages. As with any predetermined text, the content is controlled, which simplifies comparison of different readers.

For the second passage Gaudio chose an excerpt from *A Torch Song Trilogy* by Harvey Fierstein. The text used for the second recording was as follows:

> “Good. ‘Cause it’s all your fault. Couldn’t take no for an answer, you just had to come up here. Why couldn’t I just keep my trap shut? I’ve always thought of myself as a kind person. Not saintly but generously thoughtful (in a bitchy sort of way). But since being here I have said nothing that hasn’t hurt someone: you, Ed, Laurel, myself. Well, she asked for it. She begged for it . . . and boy did I give it to her. I was brilliant. Point after pointless point I proved beyond a shadow of a doubt that Ed has no idea that she even exists. That, to him, she’s simply living proof of his normality.”

This particular monologue gives the reader the option of charging his speech with a bit of emotion. I initially hesitated to use such a passage because the character in the play is a gay male. However, as Gaudio notes, the content is not specifically gay, though "the passage did, however, involve the speaker's use of the word bitchy to refer to himself" (p. 43). In the end, some of the trepidation regarding the speaker's recognition of the passage was assuaged-- none of the speakers indicated familiarity with it, and so I concluded that the readers were not biased by its origin. Nevertheless, the subjects may have a feeling the passage sounds gay, but that is something I simply could not control. For this passage the readers were given the following directions: "Please read over silently the following monologue from a play. When you're done, I'll record your reading it aloud."

I initially planned to use the same two passages as Gaudio solely, but I later decided I wanted to record more levels of speech than just these two. I therefore appended two more. In order to investigate my questions about the influence of phonetic segments, I added a word list reading. I derived the segments I wanted to test from my own list, that of , and some from Bert Vaux (personal communication).[16] Crist measured a number of tokens in his comparison of the speech of the same individuals *performing* "ordinary" and

---

16 One of the additional segmental tests that we devised is derived from . Zemlin found that "[t]here are no discernable differences in the infant's and child's larynx" and that prepubescent boys employ strategies such as over-rounding rounded vowels, which lowers formants, and avoiding smiling, which raises formants, in order to sound more masculine. This led to including words with rounded vowels in the word list.

"queeny" voice speech styles.[17] As a standard, I used his list as a starting point in developing my own word list. Reading this word list then became the third recording of each of my speakers. They were given the following directions, "Please read the following list of words aloud as I record you," after which I gave further instructions to read the words quickly but in a detached manner so that word boundaries would remain intact (in order to simplify phonetic segment analysis). The following is the word list that was originally formatted one word per line (see Appendix A for the original format):

> "1. pure; 2. smack; 3. crew; 4.leaf; 5. Jack; 6. Sprat; 7. boat; 8. spam; 9. zip; 10. tat; 11.sack; 12.fat; 13. scram; 14.you're; 15.half; 16.stand; 17.flat; 18.sure; 19.cat; 20.boot; 21.feel; 22.scam; 23.frat; 24.smolder; 25.pat; 26.strand; 27.ship; 28.laugh; 29.split; 30.Tuesday; 31.sold; 32.sat"

Directions for the final speech task were not included in the packet. I wanted to elicit and record ordinary speech from each of my subjects, so I asked them the following questions: "What has been your experience with Harvard Dining Services?" and "What was your experience like in your freshman dorm?" I was interested in having a clip of thirty seconds or less for each of the recordings. Since the content of each speech act for the first three was identical, I simply clipped the first thirty seconds of each recording. However, since the

---

17 These tokens included the following: "closure and aspiration time for /p/, /t/, /k/ in the environment #_V; frication time for /s/ in the environments #_V and #_mV; frication, closure, and voice onset time for /sp/, /st/, /sk/ in the environments #_V and #_rV, as well as frication, closure, and voice onset time for /sp/ in the environment #_lV; frication time for /f/ in the environments #_V, #_lV, #_rV; and duration of /h/ and /l/ in the environment #_V" (Crist 1997).

content varied for the final speech task and one of my goals was to elicit as neutral content as possible,[18] I clipped the thirty-second portion having the most neutral word content.

3.2 Speaker Questionnaire (Appendix A)

The four recordings described above comprise the sum of recording of each speaker's voice. However, I felt that knowing something about the person behind each voice would be helpful in later analysis, so I collected more data.

---

[18] In hindsight I see I hoped unrealistically to get fairly similar responses, and so lexical choice would not be an important variable in speech judgment. My original intention was simply not feasible. In any event, the fourth recording of each speaker is valuable as a glimpse into all of the variables a listener might use to judge a speaker, regardless of varied content. As an example of the types of content found in the free response task, the following are transcriptions of the speakers judged most gay-sounding and most straight-sounding by listeners (in that order): "Harvard Dining. Today was actually a good day to ask 'cause we had chicken fingers and that's probably my favorite. Fried food, I love it. Umm. It wasn't actually as bad as I thought it would be. I mean, clearly it's not my mom's cooking, but it's fine. It's pleasant." (0404) and "Well, when I showed up to my dorm on the first day, umm, I knew I had been good friends with, uhh, I would be good friends with, uhh, my roommate Adam, 'cause we'd talked over the summer, and we sort of knew each other and we knew we were very similar. And then we showed up, and sure enough, we were good friends. And there was another, uhh, roommate there, Joseph, who I didn't think I would be as good friends with. And sure enough, we were very different" (1404). I include these examples because one could assert that it is principally the content which sways listeners to judge speakers a certain way. However, it seems that, for being the two samples judged most extremely gay and straight, the difference in content does not merit the conclusion that content is the principal reason that listeners judge the speakers the way they do even in a free response passage.

This was not unlike Gaudio's 1994 method in devising his own experiment,[19] but unlike Gaudio's oral interview, I decided to present my listeners with a written questionnaire, allowing them to respond using whichever medium was most comfortable. In answering all of the following questions, subjects were given the option to type, speak (recorded), or write (by hand) all of their answers. Also note that some of the following questions are excluded from the total analysis because of later limiting the scope of inquiry.

3.2.1 Descriptive Questions About Basic Background

Questions about the speaker's basic background and training as an actor or public speaker comprised the first page, which included the following:

> "Please answer the following questions as completely as possible. Go into as much detail as you feel comfortable (Let me know if the space provided is not sufficient.) Please use specific examples, when possible. All information is strictly confidential and will be

---

[19] Gaudio 1994 reports the following methodology: "After recording each subject's reading aloud of the two passages, the researcher took him to an office in the same building. Using the same tape recorder with its built-in microphone, the researcher conducted and recorded an approximately 45-minute interview with each subject. In the course of the interviews, the researcher asked each subject about his family and where he had grown up and about high school and college activities, such as studies, extracurriculars, socializing, dating, and so on. The researcher also asked each subject about his knowledge of or acquaintance with any gay people while in high school and college. The researcher asked the gay subjects about their own coming-out process, that is, how they came to terms with their sexuality and began to lead lives as openly gay men. The researcher asked all subjects about their ideas (including stereotypes they held or knew about) relating to gay identity, gay behavior, and gay language. He did not explain his particular interest in pitch and intonation" (pp. 43-44).

used anonymously. Also feel free to leave blank questions you wish not to answer.

1. Where were you born? How old are you?
2. Where have you lived? How long have you lived in each place?
3. What kind of high school did you attend?
4. What do/did you study at university?
5. Have you done any postgraduate work?
6. Do you consider yourself an actor? How much experience do you have as an actor? What about your family?
7. Are you involved in public speech of any sort? Debate? Do you have oratorical training? What about your family?"

I asked where the speaker was born and his places of residence, because listeners often pick up on cues that match with their perceptions of speech variety based on location. The rationale is similar for age, except, since all of the speakers are Harvard college undergraduates, the range is roughly five years. I was hoping to make class inferences by knowing the type of high school a speaker attended, and field of study could be related to speech stereotypes as well.[20]

3.2.2 Quantitative Self-Description Questions

Since the above questions are not quantitative in nature, and it is easier to compare answers made up of numerical values and set responses, I included another set of questions that would enable me to compare speaker self-judgments against listener judgments of the same person. The first page of these questions included the instructions, "Would you use any of the following to describe yourself? (Please Circle all that apply)." The rest of the page was filled

---

[20] Gaudio 1994 incorporated ratings of humanist/scientist into his system.

with the following terms, originally presented in three columns (see Appendix A for the original format):

> Academic, African, Analytical, Artistic, Asian, Athletic, Bisexual, Black, Conservative, Dependent, Dramatic, Effeminate, European, Extroverted, Fat, Feminine, Gay, Hispanic, Independent, Intelligent, Introverted, Latino, Liberal, Lower-Class, Masculine, Middle-Class, Minority, Moderate, Musical, Native American, Patriotic, Radical, Reserved, Scholar, Straight, Thin, Tone-deaf, Transgender, Transsexual, Upper-Class, White, Working-Class.

I realized that some subjects might prefer not to describe themselves in terms of these sorts of categories (to not "box themselves in"), and I might get blank response sheets returned. (A blank response sheet could also mean, however unlikely, that the subject feels comfortable describing himself in such terms but feels that none of the terms apply to him.) However, it was necessary to provide some method of systematically comparing listener judgments to how the speaker subjects would describe themselves. These terms are admittedly not the only ones a speaker could use to describe himself, but are representative of the possible ways college students see themselves. To reiterate, this page allows for a sampling of descriptive variables that are easily quantified and thus more easily compared.

Preparing for even more direct comparisons with listener judgment data, the following two pages in the survey asked speaker subjects to describe their voice using exactly the same predetermined terms that the listener subjects would use.[21] Because humans can conceive of themselves as belonging to both exclusive categories and along a sliding scale, I elicited both groups of data from the speaker subjects. "Please pick one word from each of the following pairs which best describes the quality of your voice: (Please Circle One for each)"

---

[21] See page for a description of the selection of these terms.

were the directions for this page followed by the five pairs, one on each line: “A. dramatic or plain | B. masculine or feminine | C. reserved or emotional | D. gay or straight | E. ordinary or affected.” This formulation forced a choice between polar opposites,[22] as a step toward quantifying the way in which a speaker subject sees himself.

The following page contained these directions, “Please rate the quality of your voice on the following scales (Circle the number),” with the rubric in Figure 1:

Figure

(left side descriptors:)	Extremely left side	Mostly left side	Somewhat left side	Neutral	Somewhat right side	Mostly right side	Extremely right side	(right side descriptors:)
A. Plain	1	2	3	4	5	6	7	A. Dramatic
B. Feminine	1	2	3	4	5	6	7	B. Masculine
C. Humanist	1	2	3	4	5	6	7	C. Scientist
D. Straight	1	2	3	4	5	6	7	D. Gay
E. Reserved	1	2	3	4	5	6	7	E. Emotional
F. Affected	1	2	3	4	5	6	7	F. Ordinary

This is essentially the same question as the previous page, but it allows the speaker to express graded levels when using the same terms for self-description. Comparison with (see page) show the close similarity of this rubric and the one in which listener subjects rate the speaker. As such, direct comparison of the speaker subject’s self-description and the listener subjects’ judgments is not only possible, but relatively simple in mechanical terms.

---

[22] This, of course, assumes that the subject decided to respond.

3.2.3 Descriptive Questions on Gayness and Gay-Sounding Speech

In collecting quantitative data from speaker subjects, as described above, my method differed from the method employed by Gaudio 1994 in that the entire questionnaire portion of my data collection with speaker subjects is not an oral interview. Regardless of the *modus operandi*, I collected additional descriptive data from speakers, akin to Gaudio's described line of questioning (see footnote on page).

The page that followed in the questionnaire represents a bifurcation based on previous responses. I prepared two separate sheets in order to tailor the questions to gay-identifying and non-gay-identifying subjects, and gave the appropriate sheet to each subject at this point. The gay-identifying subjects were presented with the following directions and questions:

> Please answer the following questions as completely as possible. Go into as much detail as you feel comfortable (Let me know if the space provided is not sufficient.) Please use specific examples, when possible. All information is strictly confidential and will be used anonymously. Also feel free to leave blank questions you wish not to answer.
>
> 1. When/How did you realize you were gay?
> 2. When/How did you come to accept it for yourself?
> 3. Have you come out? If so, when did you come out?
> 4. Are you out to everyone? If you are out to your family, how do they feel?
> 5. What was your coming out experience like?
> 6. Do you feel discriminated against? In what way?

The non- gay-identifying subjects were presented with the following questions (the directions were the same as above):

1. Have you ever had a gay experience or considered it? If so, what was it?
2. Is anyone in your family gay? If so, how does you family treat him or her?
3. How do you feel about same-sex attraction and love?

4. Do you feel uncomfortable around gay people? Why or why not?
5. Do you feel intimidated by gay people? Why or why not?
6. Do you or have you ever used the way someone speaks to judge their sexual orientation?

Since the concepts “gay-sounding” and “straight-sounding” do not exist in a vacuum, this line of inquiry is important in order to understand how the speakers view themselves and their environment in terms of sexuality. The thoughts that people, both gay and straight, expressed about identity and sexuality helped me contextualize their self-descriptions of voice quality. By extension, this deeper knowledge of who each speaker subject is aids the comparison with judgments of the listener subjects who only know these people by their voice.

The final page was given to all speaker subjects, and unlike the previous pages, this one asks the subjects to comment principally on gay-sounding speech in general. The questions presented are the following (the directions are the same as above):

1. Are you aware of any speech styles specifically used by gay men?
2. Are these stereotypes or do they actually exist?
3. What are attributes of this speech style?
4. Do you think you have any features in your own speech that might be considered gay- or straight-sounding?
5. Do you have gay friends? Are any of them close friends? Do they sound gay? If so, how?
6. Can you think of any situations in which they sounded particularly gay or not gay? Do you think that the speech style was consciously chosen?
7. Does their speech pattern affect (interactions in) your relationship? If so, how and in what context?
8. Can you tell someone is gay by the way they sound? If so, how?

The data from these questions is useful not only in analyzing the subject himself, but also for analyzing other subjects’ voices and also the responses of listener subjects. After compiling the answers to these questions for all of the speaker

and listener subjects, I ended up with a descriptive database of the sorts of features that are commonly held to exist in gay- and straight-sounding speech.

3.3 Preparation of the recordings

Each of the four speech acts described in Section 3.1 was recorded in a separate sound file for a total of sixty-eight clips (four readings per seventeen speakers). In order to reduce fatigue for the listener subjects, I cut all the sound clips to under thirty seconds. For the recordings of the first two speech acts, I included the speech from the beginning of speech to the end of the last phrase before approximately the thirty-second mark for the clips. The speech recordings for the word list were clipped in a similar fashion except I included speech data through the last word (the words were separated by short pauses) before the thirty-second mark. The fourth recording was the most difficult to clip, since I consciously attempted to select the most neutral-sounding excerpt of thirty seconds. I realized later the impossibility of this goal, not only because “neutral content” is so hard to define, but also because of the widely varying responses and response levels to the question(s). It is also relatively impossible to compare phonetic segments across speakers based on the recordings of the free response question since the same phonetic segments may not exist in each recording. In sum, although I selected the portion of the total from which to make the clips with certain goals in mind, in the end it was as arbitrary as having selected the first thirty seconds of each recording.

Having prepared the sixty-eight clips, all numbered for each speaker then for each clip (e. g. 0101 for speaker one, clip one; 0102 for speaker one, clip two; 0201 for speaker two, etc.), I created a folder on my computer which consisted of all the speaker recordings arranged one after another, all the clips for one speaker followed by those for the next speaker, each set in order. However,

because I wanted the listener subjects to evaluate each clip individually,[23] I decided to randomize the clips.[24] I only made minor modifications to this randomization in order to ensure two clips from the same speaker-subject were not played in immediate succession. (I played the clips for the listener-subjects by progressing through the folder one at a time in the now randomized alphabetical order.) I must add that randomization makes the task of listening to the sixty-eight clips less monotonous for the listener-subjects, but, more importantly, it adds an element of surprise in the task. This, I feel, makes the task more spontaneous, and thus, more natural, simulating the real-world condition of picking up the phone and talking to an unknown individual, which is the type of interaction this study hopes to simulate.

3.3.1 Reasons for having four recordings

---

[23] Listener subjects were not told that there were multiple clips of the same speaker, but virtually all revealed that they assumed different speakers for each clip.

[24] My randomization algorithm consisted of inserting the letters (one per clip) from "The quick brown fox jumped over the lazy dog's back!" at the beginning of the clip name (i. e. t0101, h0102, e0103, q0104, u0201, i0202, c0203, k0204).

Using read passages (excerpts from the accounting text and play), allows for maximal control over the content, as noted earlier. Gaudio cites precedent of read passages having been used in prior studies of pitch properties (; ;), "in order to control for lexical choice, topic, and length of recording."[25] In studying gay speech and its stereotypes, control of lexical choice and topic can be especially important, as points out:

> "The recognition that lesbians and gay men may have their own words is reflected in the numerous attempts at recording them in specialized dictionaries and glossaries (Westwood 1960; Strait 1961; Cory 1965; *Guild Dictionary of Homosexual Terms* 1965; Farrell 1972; Rodgers 1972; Bardis 1980; Spears 1981; Taub and Leger 1984; Dynes 1985; and Max 1988)" (p. 55)

As noted earlier, the purpose of the word list (recording number three) was to elicit specific phonetic segments in the vein of Crist 1997. Casual speech is the register most likely to give good correlations with listener judgments, since it is the type of speech on which the preconceptions of listener subjects are largely grounded. Based on , word lists are the polar opposite of casual speech in the hierarchy {Casual > Careful > Reading > Word lists/Minimal Pairs}. Given the problems with word list register, there was little hope that strong correlations could be found in the listener, though I included it in case I wanted to have data on specific phonetic segment variation.

When I took clips of the fourth recordings, trying to control for the somewhat elusively defined "neutral content," I was cognizant of the generalization made by : "The most notable characteristic of these lesbian and gay glossary dictionaires [*sic*] is the emphasis on sexual matters." Thus, I was

---

25 I agree with "lexical choice" and "topic," but length of recording would assume speakers speaking at the same speed for a given text.

looking first and foremost to avoid this subject matter and the related terms. (But, as I mentioned, it does not matter in the end what section of the recording I chose, since listener subjects can pick out marked terms from a lexical categories beyond “sexual matters.”)

3.4 Collection of data from listener subjects (Appendix B)

In a manner similar to that in the first half of this study, I sent out advertisements over e-mail lists of the Harvard College undergraduate houses to get listeners as subjects. Again, each subject made a one-hour appointment with me with the understanding that they would be participating in linguistics research. I conducted the study in the phonetics lab of the linguistics department. Upon arrival, I presented each subject with a packet containing a cover sheet of instructions and nine pages of the scoring rubric () reproduced eight times on each page.

Figure

A. Plain	1	2	3	4	5	6	7	A. Dramatic
B. Feminine	1	2	3	4	5	6	7	B. Masculine
C. Straight	1	2	3	4	5	6	7	C. Gay
D. Reserved	1	2	3	4	5	6	7	D. Emotional
E. Affected	1	2	3	4	5	6	7	E. Ordinary

What other impressions, not included on this chart, do you have of this person based on their voice (e.g. race, class, weight, education, etc.)?

This rubric is derived from the one used by Gaudio 1994, which, in turn, is based on the work of , , and . Osgood, et al.’s technique included developing a “semantic differential” for eliciting standard judgments from subjects. Such a

technique is useful because it lends itself easily to quantification. As Gaudio explains,

> 'The alternatives are designed to represent the major ways in which meanings vary within the relevant population and ideally make use of lexicalized pairs rather than simple negation (e.g., "interesting/boring" rather than "interesting/uninteresting")' (pp. 44-45).

This rubric matches well with naïve approaches to categorization that people may use unconsciously. Putting concepts like *gay* and *straight*, *masculine* and *feminine* on opposites of a scale does not necessarily reflect the approach of everyone[26], but it is an oversimplification I am willing to concede for the sake of the experiment's feasibility and comparison with Gaudio.

Emily Best (personal communication) observes that by giving my subjects predetermined words in which to describe the voices they heard, I inappropriately directed their response to the material. For example, after hearing a given clip, a listener who might not have considered the sexual orientation of the speakers would be forced to consider such a judgment when given the pairing "straight/gay" on my questionnaire. While I appreciate this argument and agree it is a possible problem with this method, I would counter that I have taken two measures to increase the extent to which the listener can present his or her natural reaction to the speech. First, listeners are informed that they may leave any criteria blank if they have no response to the given pairing. Second, I have provided the option for listeners to write any other reactions they may have to the

---

[26] devised the Bem Sex Role Inventory, which allows masculinity and femininity to have their own scales, not be opposites in a zero-sum game. This allows for high levels of both, which is often referred to as psychological androgyny. basically rejects social categorization altogether. Given this postmodern view, this experiment would be impossible, so I have chosen to set aside this line of thought for the sake of simplicity.

speech that are not included within the scoring rubric. I concede that this still does not fully mitigate the influence of the given words on the rubric, but it does allow for using them in the freest way possible for the listener subject. I am not necessarily studying every response that a listener could have; I am, however, investigating to what extent listeners can make a reliable distinction along a certain dimension, i. e. , the criteria upon which they rely to make their judgments must originate from somewhere. I must also add that a main part of this study is to reproduce Gaudio's methodology in order to verify his results.

The pairings "straight/gay," "reserved/emotional," and "affected/ordinary" are taken directly from Gaudio's "Rating Scale," and "feminine/masculine" is derived from it.[27] I replaced "effeminate" with "feminine," as I believe "effeminate" is a word with added negative connotations but denoting much the same as "feminine" while "Feminine" is more of a neutral term than "effeminate," and therefore, more appropriate for this study as the opposite for "masculine." I added "plain/dramatic" not only because of the included dramatic passage, but because "dramatic" is a word often used to describe speech (particularly gay-sounding speech). Although I realize these pairs are neither the only possibilities, nor necessarily the best, they were used here because previous researchers employed equivalent measures.

> The cover sheet of the packet included the following instructions:
> I'm going to play for you a series of recordings of spoken texts. You will hear the voices of different men reading one of the three texts

---

[27] Gaudio submits that these pairs "represent stereotypes that are commonly held about gay men in United States society and that emerged in the interviews that the researcher conducted with the speaker-subjects in this study" (p. 45). The responses to the questionnaire in my study revealed similar stereotypes.

> in front of you or responding to an open-ended question. Please take a moment to familiarize yourself with the texts.

These instructions were followed by the three texts. Some might take issue with my giving the listener subjects the texts they would hear because this unnecessarily connects what should be an entirely oral exercise to written language. This was done to avoid having my subjects listen to practice segments, as was the method of Gaudio. Listening to practice recordings of the texts and judging them on the rubric predisposes listeners to a “standard” reading by which successive readings are compared. This would be more disruptive to collecting unbiased data than simply having the subjects read the texts. In sum, by reading the texts, the listener subjects were prepared for the coming speech clips, but had no previous (external) sound of the text in their minds with which to compare while making their judgments.

> After the texts (still on the same cover page) followed these instructions: As you begin to formulate an impression of the speakers, please rate them accordingly on the scales provided. You do not need to wait until the speaker is finished before you begin to rate him. When you have finished with the scales, answer the question below them. In the interest of time, please limit your response to the final question to only a few words.

There are some important points about these instructions. First, the listeners are notified at the outset that the recordings they will be listening to are only of men. In earlier trials of this study, omission of this fact led listener subjects to rate the speaker subjects unevenly on the feminine-masculine scale as they were unaware of the lack of female speakers until well into the experiment and were waiting for female voices to use the “very feminine” ratings. It is also important to note that listeners did not need to listen to the complete clip to begin their response. The above instructions are predicated on giving the listener a sense of

urgency. The intention is for subjects to record their first impression and move on, not taking extra time to rethink their judgments. In testing immediate responses of short segments, I follow the work of , who find that dialect identification is possible with the single word “hello.” That is, the listener begins to form a perception of the disembodied speaker with the first word of the speaker’s utterance.

Following this cover sheet were the copies of the scoring rubric discussed above. I played the sixty-eight randomized clips for the listener subjects, and they recorded their judgments. When finished, I gave the subjects an exit survey with many of the same questions as on the first and last pages of the speaker subject questionnaire:

> Please answer the following questions as completely as possible. Go into as much detail as you feel comfortable (Let me know if the space provided is not sufficient.) Please use specific examples, when possible. All information is strictly confidential and will be used anonymously. Also feel free to leave blank questions you wish not to answer.
>
> 1. Where were you born? What is your sex? How old are you?
> 2. Where have you lived? How long have you lived in each place?
> 3. What kind of high school did you attend?
> 4. What do/did you study at university?
> 5. Have you done any postgraduate work?
> 6. Do you consider yourself an actor? How much experience do you have as an actor? Are you involved in public speech of any sort? Debate? Do you have oratorical training? What about your family?
> 7. Are you aware of any speech styles specifically used by gay men?
> 8. Are these stereotypes or do they actually exist?
> 9. What are attributes of this speech style?
> 10. Do you think you have any features in your own speech that might be considered gay- or straight-sounding?
> 11. Do you have gay friends? Are any of them close friends? Do they sound gay? If so, how?

12. Can you think of any situations in which they sounded particularly gay or not gay? Do you think that the speech style was consciously chosen?
13. Does their speech pattern affect (interactions in) your relationship? If so, how and in what context?
14. Can you tell someone is gay by the way they sound? If so, how?
15. How did you decide where a speaker would fall on the gay-straight scale?

The important difference to note is the final question, number fifteen. Though the question requires making a generalization based on a large and varied number of clips (sixty-eight), knowing subjects' own impressions of their judgment mechanism offers some indication as to why they rated speakers voices the way they did. This data was used to verify Gaudio's claims that listeners were probably using pitch related cues to make their judgments.

4 Results (Appendix C)

In order to facilitate comparison with Gaudio's findings (p. 47-53), I will report mine in the same order (Sections 4.1 through 4.3). I measured average pitch, maximum pitch, minimum pitch, pitch contours, and speed of utterance using Praat (version 4.0.29) and tabulated the results in Microsoft Excel X. Excel also provided the basic statistical analysis tools for calculating regression and ANOVA. The complete statistical analyses are provided in Appendices D through K.

4.1 Listener Perception Ratings (Appendix D)

The first hypothesis to test is that listeners can accurately guess the sexual orientation of the speakers by their voice.[28]

This hypothesis is based on the claim that this is so in Gaudio's results. I never asked speaker subjects to rate their sexual orientation on a seven-point scale, but rather in the context of words used to describe themselves (see page).

Table : Speakers' Self-Described Sexual Orientation (left) and Mean Listener Ratings (right)for all speech acts

Table 1 shows the mean listeners' rating of speakers on the gay-straight scale for each of the four speech acts with an additional column containing the

---

[28] see Gaudio: "Listener Perception Ratings," Tables 1 and 2.

averages of the four. The sexual orientation column lists the orientation the speaker subjects identified on the self-description word list page of the questionnaire (see page); “s” indicates that the subject circled “straight,” “g” indicates “gay,” and “b” indicates “bisexual.” Two letters means that the subject circles two sexual orientation description terms. Two subjects, namely speaker eight and speaker eleven, marked nothing on that page, so I added the orientation that seemed to fit best, given their responses to questions they answered later in the survey (see page).

The simplest way to test if listeners were accurate on the straight-gay scale (making the temporary assumption that the terms gay and straight are a black and white pair), is to evaluate whether or not the mean listener ratings for each speaker and each speech act match the same side of the scale. That is, ratings between one and four will all be taken as a “straight” rating, and all ratings between four and seven will all be taken as a “gay” rating. (For numerical simplicity, the speaker subjects who described themselves as bisexual or by using more than one term are not included in this tally[29].) Given those restrictions, listener ratings and speakers’ identified sexual orientation matched

---

[29] Since this data does deserve treatment as much as any other, I offer a descriptive analysis. Both of the speakers who identified as both gay and bisexual match receive very neutral (all within 1 of the 4-neutral rating) rating means for all of the speech acts. (If bisexuality were described numerically in this schema, it would be the number four.) The one speaker who identified as bisexual received very neutral ratings for the accounting passage and word list, but fairly “gay” ratings for the play monologue and free response.

64% (accounting), 71% (play), 64% (word list), 78% (free), and 71% (all) of the time.[30]

Gaudio's results for this test have overwhelming matching (94% accuracy), whereas the results in this study show medium to high levels of matching.[31] Further inspection reveals a number of interesting points. First, all the means for a given speaker fall on the same side 78% of the time. But further, in the three cases (22%) that all of the means do not all fall on the same side, three of the four do. (There are no 50-50 split cases.) Therefore there seems to be a high degree of listener rating agreement as to whether a speaker falls on the straight or gay side of the divide. Second, for three speakers (#8, #10, and #11), all of the listener means fall on the side not matching the speakers' sexual orientation (in all three cases, straight). This may be some evidence against pure gay performativity. That is, if gay speech were only consciously performed, these speakers must either not be aware how their voice is perceived or are not able to change their speech style, or both. These three speakers are examples of gay-sounding, straight-identifying people (just as speaker #16 is a slightly straight-sounding, gay-identifying person for three of the four speech acts). Figure 3 makes these generalizations even more apparent: notice how clustered the points for each speaker are in general, with more than one stray point being rare.

---

[30] Note that these figures are all higher (71% for all passages averaged together) than Rogers and Smyth's 62% matching accuracy in listener identification of gay men (see page). Therefore the data in this test is supportive of their claims as well.

[31] Gaudio also claims that "a simple rank test indicates that such a distribution has a 1/70, or less than 2%, chance of occurring randomly" (p. 47). It is unclear what sort of a test a "simple rank test" is and therefore how he calculated that distribution. For that reason, it is not included here, and a simpler arithmetic comparison of percentages is given.

The implications of this data are that the type of speech act is not a crucial influencing factor in listener judgments. In other words, for many speakers, it does not matter what they say; listeners make similar gay-straight judgments regardless of the speech act.

Figure

Table 2 shows the results of analysis of variance (ANOVA) tests, which were used to determine the extent to which the identified sexual orientation of the speaker correlates with the listener ratings on the straight-gay scale. In order to perform this test, sexual orientation identifications (see page) had to be converted from words into a numerical format. The conversion scheme is given in

.

Table Correlation of Sexual Orientation and Listener Ratings for all speech acts[32]

	straight-gay	
Accounting	*r*=	**0.486**
	p=	**0.048**
Play	*r*=	**0.545**
	p=	**0.024**
Word List	*r*=	0.265
	p=	0.304
Free	*r*=	**0.690**
	p=	**0.002**
All	*r*=	**0.559**
	p=	**0.020**

32 In the comparable table in Gaudio, he also includes F values (the F value is a ratio of sample variances). F values can be ignored for our purposes; they are only important here because the *p* values are derived from them, but, Excel automatically calculates the *p* values.

On all of the summary charts, bold numerical values indicate results that are statistically significant at the level $p<0.05$ while bold row names indicate results at a lower level of statistical significance, $p<0.10$.

Table Conversion Chart of

Sexual Orientation Descriptors to Numerical Values

Word:	straight	straight & bisexual	bisexual	gay & bisexual	gay
Numerical Value:	1	2.5	4	5.5	7

For the purposes of this study, the *p* values basically give the probability that the results occurred by chance and had nothing to do with the experiment. Therefore, in hypothesis testing, low *p* values are desirable (they range from 0 to 1). In other words, *p* values tell us whether or not the two variables correlate in a statistically significant manner. The values $p<0.05$ are a commonly used range in determining statistical significance (this level means being 95% sure that the correlation is not due to chance). Given that parameter, all correlations except for the word list vs. straight-gay are statistically significant. The other value, r, is the correlation coefficient. This value can tell us how well one set of data can predict another set, in this case how well the self-described sexual orientation of a speaker predicts what score the listener will rate them on the straight-gay scale. An *r* value of 1 would tell us that one data set perfectly predicts the other, whereas an *r* value of 0 tells us the relationship is totally random. In this case, the range of *r* values from 0.486 to 0.690 (excluding the word list which is not statistically significant) tells us that the sexual orientation of the speaker is a good predictor of listener judgments on the straight-gay scale. In sum, for speech acts besides the word list, the statistical analysis supports the hypothesis that listeners can guess the sexual orientation of the speakers based on their voice.

4.2 Pitch Range (Appendices E and F)

Gross pitch range is calculated by subtracting the minimum pitch value from the maximum for the entire clip.[33]

shows the *r* and *p* values of the correlation of pitch values and listener ratings on the straight-gay scale[34]. None of the values show the correlation to be statistically significant (at the level set before, $p<0.05$). Given the lack of statistical significance, it hardly matters that the correlations (*r* values) are extremely weak for three of the four speech acts and somewhat for the play monologue. These data suggest that gross pitch range is not a significant variable used by speakers judging a voice straight- or gay-sounding.

Table Correlation of Gross Pitch Range

and Listener Ratings for all speech acts

In order to get a possibly even truer picture of the pitch range of a speaker's voice, I (following Gaudio) also used the middle 95% pitch values to make another comparison of pitch range to listener judgment. (Cutting off 2.5% of the pitch values at the minimum and maximum ends of the pitch range helps to compensate for a random extreme pitch drop or rise that is not part of the characteristic pitch range for the entire clip. In effect, this analysis, as opposed to considering all values, mitigates the effect of outliers.) shows the results of this

---

[33] All pitch calculations are made in Praat at a resolution of 0.01s time steps.

[34] F values were omitted as not being necessary for this summary. The values can be found in Appendix D.

analysis. Significance (and good correlation) is lacking for all speech acts except for the free response.[35]

In the free response recording, the mid-95% pitch range correlates very well with the listener ratings. Because of this radical change from the analysis when considering the entire pitch range, we can conclude that the 95% pitch range data can be valuable in assessing listener ratings. It seems that outliers were, in fact, greatly affecting the correlation, so that the relationship between pitch range and listener response was unclear.

Table Correlation of Mid-95% Pitch Range and Listener Ratings for all speech acts

In order to understand this difference, one analysis suggests that during the free response, speakers are using a speech register closest to their everyday speaking voice. The importance of this fact is that listeners are accustomed to judging voices with conversational speech register making up the vast majority of the situations. The accounting passage and play monologue, on the other hand, are both variations of a “reading voice” register. The word list, while requiring the speaker to read, yields a different register, as can be seen in (p.84). Therefore, in this paradigm, the free response clips represent a very different type of speech

---

[35] For the corresponding pitch range tests, Gaudio reports a lack of significance at the $p<0.05$ level. However, at the $p<0.10$ level, which he terms “suggestive” (this is not a standard label for this level in statistics), he has supportive results for the accounting passage. Correlation is pretty good with r values of 0.6201 (gross pitch range) and 0.6535 (mid-95% pitch range). Even at Gaudio’s “suggestive” level, my data for the accounting passage does not support the hypothesis that pitch range can predict listener rating on the straight-gay scale.

register. The pitch range used in reading is likely to be very small and have less interspeaker variation than that of regular conversational speech. While content is a major variable that can influence listener judgments in the free response clips, unlike in the read passages, when comparing the pitch ranges of gross pitch range versus mid-95% of pitch range, this variable is not relevant (that is, the measurement of pitch range is different but the clips themselves are identical). Therefore, for the free response clips, it seems that mid-95% of pitch range is a good predictor of listener judgments on the gay-straight scale.

4.3 Pitch Variability (Appendices G and H)

Following Gaudio (p. 50), I employ two ways of calculating variability. The first method is based on , which uses values of calculated variance and standard deviation. The claim is that these statistical results are a way to represent pitch variation. shows the *p* and *r* values for comparison of statistical variance and standard deviation for the entire pitch range[36]

against the listener judgments on the straight-gay scale. The *p* values are significant for the free response passage only for both variance and standard deviation. In addition, the high *r* values for the free response passage show a good correlation between the data sets. Therefore, for the free response passage

---

[36] While Gaudio and Aronovitch calculated the corresponding values for the mid-95% pitch range, as seen above, I was unable to do so using the Praat sound analysis software.

measures of variance and standard deviation for variability are good predictors of listener response on the straight-gay scale.

Table Correlation of variance (left), standard deviation (right) and Listener Ratings for all speech acts

	straight-gay		straight-gay
Accounting	*r*= 0.129	Accounting	*r*= 0.194
	p= 0.622		*p*= 0.456
Play	*r*= 0.333	Play	*r*= 0.361
	p= 0.191		*p*= 0.155
Word List	*r*= 0.032	Word List	*r*= 0.014
	p= 0.902		*p*= 0.957
Free	*r*= **0.569**	Free	*r*= **0.531**
	p= **0.017**		*p*= **0.028**

The second method of calculating pitch variability is based on [37]. It involves measuring the change in the pitch every 0.01 seconds and calculating the mean of those measurements.

shows the *r* and *p* values for the comparison of the mean pitch change and the listener ratings. None of the *p* values are less than 0.05, so there is no statistical significance at that threshold. Therefore, this measurement of pitch variability is not likely to be a good predictor of listener judgments on the straight-gay scale.

Table Correlation of mean change in pitch and Listener Ratings for all speech acts

[37] Eady and Gaudio actually used two methods to calculate pitch variability. I do not use the second method here because it seems superfluous and is extremely time-consuming.

	straight-gay	
Accounting	*r*=	0.103
	p=	0.694
Play	*r*=	0.390
	p=	0.122
Word List	*r*=	0.055
	p=	0.835
Free	*r*=	0.425
	p=	0.089

4.4 Overall Mean Pitch (Appendix I)[38]

shows the relationship of mean pitch over the entire clip against listener judgments on the straight-gay scale. As with the measure of mid-95% of pitch range, only the free response shows statistical significance. Its corresponding *r* value shows the correlation is moderately strong as well.

Table Correlation of mean pitch and Listener Ratings for all speech acts

4.5 Speed of Utterance (Appendix J)

I measured the speed of the entire utterance for the three controlled-content clips. Because of irregular pauses and verbal hedges, measuring speed of utterance is too difficult for the free response recordings. In spite of the fact that word lists do not provide a good test of speed of utterance, word list data is included here for the sake of comparison. Table 9 shows that there is poor

---

38 The previous results sections replicate Gaudio's procedure while the methods of analysis for the remainder of the results sections are mine.

correlation and lack of statistical significance between speed of utterance and listener ratings on the straight-gay scale. Therefore, speed of utterance is not a likely predictor of listener ratings on a straight-gay scale.

Table Correlation of mean pitch and Listener Ratings for all but free response speech acts

	straight-gay	
Accounting	*r*=	0.248
	p=	0.337
Play	*r*=	0.083
	p=	0.751
Word List	*r*=	0.246
	p=	0.342

4.6 Interpassage Rating differences (Appendix K)

This analysis tests the hypothesis that speakers sound more neutral (closer to a *4* rating) when reading than when speaking; it would follow that speakers who are rated more gay-sounding when reading would be rated even more gay sounding (a rating of more than *4*) when speaking extemporaneously, and speakers who are rated more straight-sounding (a rating of less than *4*) when reading rated even more straight sounding when speaking extemporaneously. In answering this question, I compare the data from the listener judgments of the two reading passages (accounting and play monologue) against those of the free response speech. The word list is being excluded from

this comparison, because word list register does not pattern with reading register and is far removed from casual speech register (see and). shows the results of this analysis.

Table Correlations of Listener Ratings of Accounting Passage and Play Monologue Readings to Free Response Speech Act Listener Ratings with Separation of Speakers for the Accounting Passage and Play Monologue Based on Gay- and Straight-sounding Ratings

	straight-gay	
Gay-sounding Accounting	*r*=	0.694
	p=	0.084
Straight-sounding Accounting	*r*=	**0.798**
	p=	**0.006**
Gay-sounding Play	*r*=	**0.920**
	p=	**0.0002**
Straight-sounding Play	*r*=	**0.928**
	p=	**0.0001**

All of the *p* values show “suggestive” (Gaudio’s term) statistical significance ($p<0.10$), the bottom three showing significance well beyond the standard I have been using ($p<0.05$). All show high very levels of correlation, as well, the bottom two (in Table 10) to an extraordinary level. Overall, the results show good support for the hypothesis that speakers sound more gay or straight when speaking extemporaneously than they do when reading. This suggests that we should pay close attention to smaller indications of correlation with listener judgments in read passages because they can signal larger correlations for the speaker’s natural voice (that is, casual speech register).

It is also important to note that the correlations are better for the play monologue than the accounting passage. This suggests that, even with controlled written content, speakers may opt to employ features of a different speech register (i.e. casual speech) depending on the reading task. Regarding the two reading passages in this study, the correlation values in Table 10 provide fairly persuasive evidence that speakers use more features along the lines of

their casual speech register when reading the play monologue than when reading the accounting passage.

5 Discussion

As I have mentioned, I am disappointed with Gaudio 1994 and Crist's 1997 use of statistical inference with such small sample sizes. Based on Gaudio's results, later literature (e. g. Jacobs 1996 and Crist 1997) cites conclusions that listeners can pick out the voices belonging to gay and straight individuals. I am concerned that statements of (supposed) fact are based on such unreliable statistics. A minor criticism is that Gaudio's statistical result reporting is sometimes unclear in regard to method and presentation. But the larger problem is his very small data set (eight speakers, thirteen listeners), which can only lead to risky conclusions. In spite of these problematic assertions, I performed identical statistical analysis on my data so that it can be compared to the earlier experiments. My data set, while much larger than any of the aforementioned studies, is still too small to guarantee reliable statistical inferences. (For statistical analyses of any kind, the larger the data set, the better. Therefore, my data and analysis is the most reliable yet on this topic.) Also, Gaudio does not mention the problem that the statistical analyses he performed presuppose the data having a statistically normal distribution,[39] which this data, based on ratings in a small number of possible categories (here, 1-7), is very likely to not have. In this sort of situation, other statistical methods can be employed to better estimate the correlation, for example, a rank correlation, like Spearman's, can be used (see , p. 169-174).

---

[39] Data distribution is plotted graphically as a histogram (See Appendix C for examples). A normal distribution results from a data set for which the mean, median, and mode are equal, informally known as a "bell" curve (the "Histogram for 1101" is a fairly good example).

Because the aim of this study is not to connect the listener ratings of a speaker with the speaker's actual sexual orientation identification, but rather to understand which variables play a role in listener subjects' ratings, I originally did not want to present the data comparison of listener ratings on the straight-gay scale to the self-described sexual orientations of the speakers. Because I believe there are a number of straight-identifying speakers who fall into the gay-sounding stereotype and vice versa, I thought this query to be on the lower end of importance. I include it not only because I think readers will be interested in the result, but also because it was part of Gaudio's study and is quoted in later literature. This follows my goal of providing verification of previous data and analyses.

One of Gaudio's conclusions not entirely supported by the data here is that he thought that pitch range and variability were not the main factors in predicting listener judgments. Because of those findings, he suggests focusing instead on analysis of phonetic segments, though small sample size may be the culprit. Another problem with the experimental setup that Gaudio understresses is that it relies only on a speaker's reading voice. Restriction to one register, or even variations within a register, does not allow for the sort of pitch range and variability that extemporaneous speech does. That is a more likely reason that he didn't get better results, not necessarily his assertion that he wasn't looking in the right place. The results of the analysis in Section 4.6, Interpassage Rating Differences, seem to support the hypothesis to a high degree: the hypothesis is that speakers sound more neutral when reading than when speaking, that speakers who are rated more gay-sounding when reading would be rated even more gay sounding when speaking extemporaneously, and speakers who are rated more straight-sounding when reading rated even more straight sounding

when speaking extemporaneously. The implications of these data for further research in the area, if not sociolinguistics in general, is huge. If we take the speech register of the free response recordings to be relatively close to natural speech register, we can assume that the listener judgments are similar to how the listener would perceive the speaker's voice in a natural setting. That is, whatever cues the listeners are accustomed to using in judging (natural) speech should be most pronounced in their judgments of the free response. This claim is supported by the data in Tables 5, 6, and 8 (and to a lesser extent, Table 7), where the free response passage is the only one to show a statistically significant good correlation with the various measures of pitch range and variability. Table 10 shows that the more neutral judgments from reading passages are nevertheless often on the same track as the judgments to the free response. Table 10 even shows that there is a significant difference in reliability of the data relationship and correlation between the accounting passage and play monologue (the judgments on the play monologue being a far better predictor of the judgments of the free response). What all this suggests to me is that using reading passages as the sole source of data is risky. When listener judgments are overall closer to neutral and show less contrast, the cues they are using are not as clear. Therefore, it is imperative to include at least one free response recording in order to calibrate the judgments on speech acts whose speech register is more removed from natural speech. This discovery lends a measure of validation to this study as, in part, an improved, expanded version of Gaudio 1994.

6 Conclusions and Further Research

One of the important related questions that is not addressed in this study, nor can it be addressed in an experimental study, is the genesis of marked gay speech. Bert Vaux (personal communication) shared my bewilderment that friends and acquaintances from every type of background *can* manifest one or more of the characteristics that listeners define as gay-sounding speech.[40] How does this happen, given that many individuals who exhibit these identified features report they had no model of gay-sounding speech when growing up? (This does not even begin to address the sticky issue of performativity [see Zwicky 1997], which here refers to the claim that gay-sounding speakers *choose* to sound the way they are perceived to sound.) And even more perplexing, if gay-sounding speech is "performed,"[41] how are there individuals who are unaware of their gay-sounding speech patterns, speaking that way regardless of their sexual orientation. I examined the theories of speech communities in to try to understand how these "deviant" speakers might be possible, but no evidence is given—only to the contrary: the jumble of idiolects in speech community should produce a similar jumble in subsequent generations, not produce individuals with such marked patterns. For those readers who are unconvinced of the cognitive strangeness of the previous questions, a real mind-boggler that unites some of the previous questions is the following: how can individuals from separate and even isolated communities develop marked speech patterns in the context of their communities' speech patterns, but which share features? I wish the answers

---

[40] I draw upon the answers to the Listener Survey, especially the answer to question fifteen, to understand this data set (see Appendix B).

[41] I realize this is a gross oversimplification of the claims about gay speech performativity.

suggested by this study were anywhere close to those needed for these infinitely more complex questions.

One thing to conclude from this involves the stereotypes that are commonly held and the prejudices that exist in the minds of everyone. I say this because some of the listener subjects later commented to me that participating in this study made them rethink their prejudices about gay speech, even ones that they would rather not have, but tacitly keep in their minds. But the fact that listener subjects can participate in such a study and rate clips without reservation speaks to the fact that people do have these generalizations and find them useful, using them without much analytic worry even though their usage helps to perpetuate stereotypes that they may not consciously endorse.

Although Gaudio's experimental setup controls for content entirely, it also, in a way, controls the variables he wishes to analyze, those of pitch. That is, both of the speech acts in the experiment are reading passages. This restricts the output to "reading voice." I appreciate his wish to get different levels by choosing a boring accounting textbook passage and a dramatic monologue from a play, but many of my subjects treated the two tasks as if they were identical: *read* a passage, *read* a passage. This is why I made the two additions to the speaker recordings corpus. Including a word list recording allows for a different level of voice, but in this case as well, the voice is controlled. The word lists turns out to have an even greater problem in that the voice register for them is even further from extemporaneous speech than the normal reading voice register. (As has been mentioned before, this helps explain why across-the-board listener judgments for word list speech clips are less predictable.) The free response clips are only minimally controlled for content (by having asked the same questions) and have a similar register to normal speech that we hear the vast

majority time in everyday life. It makes sense, then, that in general, the listener judgments of the free response clips are most predictable across a number of different methods of measurement.

A cursory glance at the Appendices shows how much data analysis is involved in working with only a sliver of the collected data. Appendix C contains the complete set of listener judgment data. I have only rigorously analyzed one column of the six that appear on every page (judgments on the straight gay scale). There are also the data collected in the speaker survey (see Appendix A), almost none of which is quantitative. There is also a great deal of data in the listener survey (Appendix B) outside of the scoring rubric; again, almost none of this data is quantitative. That I have principally focused on the one column in Appendix C is what I mean by a sliver. That is not to say that I believe this study incomplete as it stands; it replicates and expands on the seminal work in the field, answering new questions and raising even more.

To clarify, I see nothing wrong with sociolinguistic studies with small sample sizes. My trouble is the use of statistical analysis on these small data sets to make inferences. Especially because this field is so new, I would hope that those who have time and resources for only small studies will instead focus on descriptive, but still linguistic, analyses. Once we have a better understanding, a larger (or even existent) descriptive database, then large, highly-focused studies with extensive statistical analysis will be more valuable. Because my study is multi-faceted, an extensive descriptive study would be possible as well as a rather substantial statistical analysis (although, ideally studies will be at least twice as large as mine in the future in order to surpass the fifty subject minimum criterion). It all depends on how you slice the enormous data set.

This presentation of the material may seem to jump the gun by using extensive statistical analysis on gross measurements, but, as I have mentioned, this paper allows direct comparison with earlier studies as well as expanding the knowledge in the area. Specifically I have shown that testing various registers is essential to understanding the outlined sociolinguistic phenomena because faint correlations in one register (here, reading voice) can easily be overlooked if data from other registers (here, casual speech) showing stronger correlations were not present. I also collected data supporting the hypothesis that listeners can by and large identify sexual orientation accurately and consistently. Given that result, the drive toward finding the reason for that correlation is even stronger.

In the future I hope to analyze the totality of the data I collected. I would like to test the hypothesis that gay-sounding correlates with feminine-sounding, which in turn correlates with affected-sounding, which in turn correlates with dramatic-sounding, which in turn correlates with emotional-sounding. There is also the more difficult test of trying to correlate speakers' responses to questions on the questionnaire with listener judgments, though the questions with set responses will ease performing these comparisons. In addition, there is a wealth of data on phonetic segments and other pitch variables, many of which can be tested for correlation with listener judgments (it will be important to try to determine what features are more or less pronounced when comparing different speech registers, natural connected speech and reading registers, for example). These investigations are all important to furthering the understanding of the gay-sounding speech stereotype that seems to exist in American English. If that methodology proves fruitful, I will perform this experiment in other languages in order to test these hypotheses within those languages and cross-culturally in

order to begin to understand how humans approach sexual orientation, regardless of language or culture.

References