Introduction

What is Gay-sounding (GS) speech? This question is not uncontroversially proper to linguistics. As late as 1994, Gaudio noted, "Indeed, with the exception of some studies of gay men's lexicon, the 'proper work of linguists' is barely being done at all."

Why should this be the case? A conspiracy of misconceptions and prejudices work to impede a rigorous, empirical research programme, and to limit the number of commentators available to the programme most recently reinvigorated by Gaudio. One such prejudice, which I myself had internalised until recently, is that GS speech must emerge late in adolescence, and is therefore as marginal to "serious" linguistic research as second-language acquisition. Smyth and Rogers (2002) have successfully countered this prejudice with their finding that GS speech often emerges in early childhood.

Another prejudice is that there are real languages – indigenous, signed, regional, other minority variants – that are marginalised more grievously than GS speech. This claim is constituted within a social context that often views GS speech (and, indeed, homosexuality itself) as a choice, something adopted late in life, possibly to accrue benefits such as membership within a prestigious,
upwardly mobile, conspicuously wealthy White male Gay culture. It is my position as a Queer man of mixed ancestry and working-class background that this prejudice inexcusably negates the experience of GS boys of multiple classes and ethnicities; ignores the stigmas and social constraints that even privileged GS men must face; and perpetuates falsehoods. Furthermore, the refusal of our discipline to subject GS speech to a serious, rigorous and broad set of analyses perpetuates the same misdeeds.

This paper, therefore, belongs to a larger genre of linguistic defiance – direct confrontation of multiple types of marginalisation, from within and without, of one's linguistic forms and variants. Jordan (1988), Anzaldúa (1990) and Solomon (1994) – writing on African-American Vernacular English, the multilingualism of Texan Chicana/o communities, and on American Sign Language respectively – actively reject stigmas attached to their language communities, just as I will raise gaping lacunae in our discipline's attention to GS that can only be attributed to such stigmas and prejudices. These stigmas and prejudices act upon GS boys and men, and speakers of all marginalised linguistic variants, to foreclose the possibility of solidarity, mutual support, organised resistance, self-esteem and identification of fellow in-group members, and even to impede cognitive development. It is important for the discipline of linguistics to do its part in demolishing these effects.

The purpose of this paper, and the study it presents, is to consider the nature of GS speech as a linguistic entity or entities, to argue for the latter characterisation, and to demand that our methodologies follow from this. The study will be presented in §1, followed by a discussion of the
consequences for the theory and methodologies of GS speech in §2. I argue that GS speech cannot be imagined as a single, discrete linguistic system, but rather as a set of varyingly salient divergences from Straight-sounding (SS) speech.

1.1 Methodological innovations

Since I focus critically on the ontology of GS speech and on the shortcomings of previous methodologies, I have adopted a methodological template heavily based on previous studies – notably Jacobs, Smyth & Rogers (2001), and Crist (1997) – with only small but salient innovative contributions. Methodological components I contribute include (1) an approach combining qualitative microanalysis with quantification (c.f. Mendoza-Denton 2007, Johnson 1997, Schegloff 1987), paying close attention to the attributes of each data point before averaging them out; (2) my own subjective judgments of the data, as a participant-observer and in-group member, and the subjective judgments of the participants themselves; and (3) heightened reflexivity.

Since self-reporting is often looked askance at in "serious" sociolinguistics, it is necessary to explain what, exactly, subjective judgments contribute to the analysis. Neither linguists nor experts on Gay issues outside the discipline sufficiently understand what GS individuals are indexing, or how. Therefore, we are at the mercy of in-group members' subjective statements in order to know in the first place what variables to measure and control. It is our role as participant observers to investigate subjective statements with a critical eye, of course, but this does not entail ignoring them!

Average pitch has received far too much attention in the literature on GS speech. Since
Lerman & Damsté (1969), this thread has continued through Gaudio (1994) up to Rogers, Jacobs & Smyth (2001), who note (italics mine), "[a] prevailing belief concerning gay male speech is that it mirrors the patterns of a stereotypical high-pitched women's voice". This folk knowledge has thankfully led to investigation of other pitch phenomena – patterns of pitch, for example – than average pitch, yet average pitch has remained central to this work despite repeated failure to find significant differences between GS and SS speech.

Attention to subjective data could have mitigated or controlled this tendency in the research. My study participants could name examples of men who "sound higher" than other men do, but also know many men who sound Gay despite having a low voice. But three of them, without prompting, named an "up-and-down" melody as a significant factor, thus (1) indicating at least some impressionistic ability to distinguish between pitch dynamism and average pitch, and (2) pointing in the direction of a pitch phenomenon that has not been sufficiently researched, despite Levon's (2006) and Gaudio's (1994) explicit mention of this research direction.

1.2 Experimental design

Participants were recruited by flyers posted at the University of Minnesota's Minneapolis campus, and from within my network of friends and acquaintances. They were aware that gender and sexual orientation, practice and identity were components of the participant selection process. Twelve speakers participated, of which three were removed due to poor recording quality and background noise. Of the nine participants included in the data, all spoke North American English.

---

3 This experiment was approved by the Institutional Review Board of the University of Minnesota within the rubric of the course “Language Borderlands,” under the tutelage of Amy Sheldon, PhD.
as their first language. Their ages ranged between 21 and 27. Three were enrolled in graduate programmes at the University of Minnesota; one was an undergraduate; and the other five were non-students. Seven were principally of European, one of African, and one of Asian ancestry.

As in Jacobs, Smyth & Rogers (2001), I planned upon both a speaking and a listening task; however, only the former is reported upon here. Also like Jacobs, Smyth & Rogers, I used a scientific text (the Rainbow Passage, designed by Fairbanks (1960) to be phonetically balanced) and a dramatic text (Crist 1997) as well as a free-response question. I asked speakers to complete each task using "a casual pace and a normal speaking style, talking just as you would talk in ordinary conversation" (henceforth, NS=natural-sounding). Then, as in Crist, I asked speakers to reread the dramatic text and "sound Gay". Unlike Crist, I did not ask them explicitly to approximate an extreme stereotype, but rather to produce a rendition they could imagine themselves rendering in an everyday conversational context with fellow Gay men where they felt comfortable sounding Gay. Finally, I asked speakers to reread the dramatic text again, but to "sound Straight" this time.

1.3 Analysis

In a striking departure from previous work, which has been informed by sociophonetics, I adopted a categorial phonological analysis of prosody, based on my assumption that hearers identify GS speech on the basis of multiple individuable features which have categorial representations in speech processing. Therefore, my analysis is necessarily more impressionistic than that of previous researchers, but provides a complementary counterpoint to those methods. In fact, the specific pitch dynamism features that I look at are very difficult to measure without a
phonological analysis, which may account for previous researchers' inattention to them. I coded all transcripts for prosodic commas and periods, upsteps and downsteps, prosodic accents, as well as a rough, relativistic estimate (aided by praat) of left- and right-edge pitch for each prosodic phrase.

1.4 Quantitative findings

Although the bulk of the findings are qualitative, I did find a couple of statistically significant corollaries. I did not expect to find any at the outset, since my data set is so small. The statistically significant variables I identify, furthermore, have been mentioned, but never tested for, by previous researchers.

Each prosodic variable showed large interspeaker variation, which is predicted by the model I propose in §2.3. Therefore, the dependent variables I analysed were the differences between their instances in the GS and the SS task (t-tests, df=8). For the full sample, the instances of (1) two downstep syllables in a row; (2) two upstep syllables in a row; (3) a downstep followed by an upstep syllable; and (4) multistep syllables were all sensitive to the GS/SS distinction. Differences were significant for (1) (p=.025), (2) (p=.1), the composite of (1-3) (p=.025) and also for (4) (p=.1). The strongest significance was in fact for the composite of (1-3), suggesting that the three patterns may co-function as a bundle of features in production of GS speech at least for speakers, indicating perhaps that speakers expect hearers to process these features as a bundle as well.

I devised a crude metric to determine, for each participant, whether he rendered the GS or the SS task more "similarly" to the NS task, and then added this metric to the independent variables.

---

4 The fourth logical possibility – a downstep preceded immediately by an upstep – was not associated with the GS/SS distinction. This is understandable in light of the fact that the up-down pattern is the most common in all speech forms and is therefore not as likely to be deployed as a sociolinguistic index due to its lesser salience.
The subsample whose NS was more similar to their GS speech (i.e., speakers who naturally sound more Gay) showed a significant difference only in the down-down pattern (p=.1, df=4). The other subsample (i.e., speakers who naturally sound more Straight) showed significant differences in all of the above features as well as an additional one, total changes in pitch direction (p=.025, df=3). Thus, it appears that men who naturally sound more Straight can or must deploy more variables more successfully in order to switch from one register to the other, while men who naturally sound more Gay may be less adept at this switching, which would be consistent with early acquisition of GS at the expense of SS speech.

1.5 Qualitative findings

In the following, I run through findings specific to individual participants or subgroups of them, many of which findings are based on transcript analysis, while others are purely subjective and derive from my judgment as participant observer and in-group member or from the judgment of the participants themselves.

Only two of the participants sounded relatively similar in all three tasks, according to the crude metric described in §1.4. For one of these two, I was unable to find significant differences in any of the performances at all; for the other, I noticed mostly non-prosodic corollaries, such as (1) the use of glides in the GS as opposed to glottal stops in the SS task to separate adjacent vowels; (2) opener and laxer production of several vowels and diphthongs in the GS task; (3) fewer pauses in the GS task; and (4) the oft-cited strident GS sibilants.

One subject produced extremely similar NS and SS renditions, while his GS performance
struck me as highly exaggerated and not a "realistic" rendition of either "authentic" or "stereotypical" GS speech. Another, on the other hand, produced very similar GS and NS renditions, which differed only in degree to which features were used, not in which features were used; and his SS performance sounded as caricatured and unrealistic as the other subject's GS had. This was remarkable, because although there was (as predicted by my model in §2.3) considerable diversity in the GS renditions in general, all of the participants except this one were able to produce convincing SS renditions – due to the necessity of possessing this repertoire in our cultural context.

Five different participants tended to open their low vowel [ɛ] as well as the diphthongs in so, going, out, late, guys in their GS task, and to round [ʊ] in their SS task. It is remarkable that all of these features co-occurred in all five of these participants.

One subject did a convincing dramatic rendering for both the GS and the SS task. His GS features were similar to his NS features, but assumed a more exaggerated prominence and frequency. Another interesting effect he used in his GS was to alternate between high-intensity, loud reading and soft reading in order to bring out the dramatic nature of certain parts of the text. Pitch and other effects were bundled together in this terribly complex technique. This one subject's style of highly dramatised rendition of stereotypes is useful enough in analysing perception cues that it could serve as a study of its own.

A few participants appeared to use phrase-final word duration to index the GS/SS distinction.
Gay sounds: A non-discrete model of Gay speech

1. phrase-final word durations for Subject 3

<table>
<thead>
<tr>
<th>Word</th>
<th>location</th>
<th>SS duration</th>
<th>GS duration</th>
<th>other notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>happened</td>
<td>line 1</td>
<td>0.44s</td>
<td>0.55s</td>
<td>most of the difference in 1ˢᵗ syllable</td>
</tr>
<tr>
<td>street</td>
<td>line 4</td>
<td>0.42s</td>
<td>0.49s</td>
<td></td>
</tr>
<tr>
<td>there</td>
<td>line 6</td>
<td>0.34s</td>
<td>0.67s</td>
<td>extreme difference!</td>
</tr>
<tr>
<td>trapped</td>
<td>after line 9</td>
<td>0.43s</td>
<td>0.68s</td>
<td>late voicing cutoff in GS /tr/</td>
</tr>
</tbody>
</table>

Single tokens of this index, such as that in *there*, or the first syllable of *happened*, could easily serve hearer's interpretation of this sample as GS in combination with other features.

Another interesting effect is general phonolexical reduction.

2. phonolexical reduction by Subject 3

“I was just sitting here studying, and it was getting pretty late, and I was going to go to bed here pretty soon.”

SS – [ʔəwəzʒəsɪrɪŋhi˞ʃtəɾɪɪŋ æ̃ɪʔwəzɡɛ̧ɾɪŋpʰɹɪɾɪʃ ænəwəzɡɔndʒoʊɾəbɛ̧dˈhiʃpʰɹɪɾɪʃun]

GS – [əʊssɪrɪŋhi˞ʃtəɾɪɪŋ ænɪʔwʌzɡɛ̧ɾɪŋpʰɹɪɾɪʃ ænəwəzɡɪnəɡəɾəbɛ̧dˈhiʃpʰɹɪɾɪʃun]

S\G – ʔ ə ʒəs o d r

The S\G line gives the segments realised in the SS performance without correspondents in the GS rendition. This participant drops the word *just* entirely in the GS task, replaces *going to* with *gonna*, collapses *I was* to a single syllable with no glottal onset, and has no medial consonant in *pretty*. There is a general impressionistic sense, less demonstrable but equally perceptible, of general underarticulation in Subject 3’s GS rendition. It is unclear what Subject 3 is indexing through this device, but it runs counter to previous claims that GS speech is "more careful" or "articulate". I believe this correlation may be an epiphenomenon of the fact that the tasks were carried out principally on a dramatic text: dramatic texts demand more emotional readings, but displays of emotion are problematic within the heteronormative masculine standard that is SS speech, so perhaps underarticulation associated with emotional speech may be associated with the GS task for this reason.
Two speakers used a gruff, hyper-creaky voice ("fry") to index Straightness; others used low creak relatively often at the ends of sentences in GS (which, of course, made pitch analysis quite difficult). Two speakers used a GS strategy that involved lengthening both vowels and consonants in particularly emphatic or dramatic words such as *please*, *screaming*, etc. These devices have low token frequencies but high salience in conversation and must be a part of a serious analysis of the nature of GS speech. But their low token frequency makes them difficult to subject to quantitative analysis, which is why previous research, with its pervading quantitative focus, has failed or neglected to analyse them.

1.6 The scientific text

Although this work dwells less on the scientific passage and the free response than on the dramatic passage, I was able to duplicate previous findings (Smyth & Rogers 2002) that there was less variation in the scientific passage\(^5\).

2.1 The discrete model of Gay speech

Since Lerman & Damsté (1969), the finding that hearers consistently identify GS voices has often been replicated. GS speech, by definition, is a category imposed by hearers upon speakers. Successful speaker self-indexing\(^6\) as GS or SS, conscious or not, demands attention to social conventions retrievable and interpretable by hearers. Depending on the hearer and the context, GS speech may bring benefits or costs. Yet GS individuals vary considerably in their attention and

---

\(^5\) Smyth & Rogers found that the only difference in listener judgments between the scientific passage and the dramatic one was that some SS speakers sounded a little more Gay. I believe that this is due to perception effects: if SS use requires active performance of the heterosexual masculine script, then listener judgment may demand such performance from men in order to be evaluated as Straight; and a dry, scientific passage, unlike a dramatic piece, offers men little chance to perform e.g. a flat, nonchalant, unengaged lack of emotional response.

\(^6\) My theoretical background on *indexicality* is derived from Silverstein (1976), i.a.
response\textsuperscript{7} to this fact. They also vary considerably in age of onset of GS features, and, as I will demonstrate, in \textit{which} features they employ.

It is evident that GS speech is not a discrete, internally consistent linguistic system. Researchers have failed to find direct correspondences between linguistic variables and hearer identification of GS speech, and this failure stems, as Zwicky (1997) argued, from the fact that GS speech (and Gayness in general) are unified only in their divergence from prescribed, stereotypical, heteronormative masculine behaviours and speech forms. The surface form of this divergence is unimportant; only the \textit{fact} of divergence is important, and hearers are variably skilled at recognising a myriad of different forms of divergence.

\textbf{2.2 Previous research}

The "failure" Zwicky mentions has followed from methodologies which (1) reproduce a stereotype linking a particular linguistic variable to GS speech; (2) elicit speech; and (3) subsequently analyse data with or without listener judgments. This sums up the state of research on "Gay speech". Absent from such setups are, i.a.: (1) attention to the non-discreteness of GS speech; (2) prior ethnographic or qualitative analysis of individual variation; and (3) critical evaluation of the malleable cultural, perceptual and other factors determining the stereotypical links in question. Very recently, other commentators (in particular, Runner 2004 and Levon 2006) have vociferously attacked these shortcomings\textsuperscript{8}.

\textsuperscript{7} This "response" may be described as \textit{passing}, which I attribute to Larsen (1929), although she did not invent the term. For discussions of the benefits and costs of passing, the reader is referred to McDonald (2000) and Kondo (1990).

\textsuperscript{8} I would urge the reader at this point not to fault other Gay researchers, however. Their methodologies have not been inherently faulty, but rather these shortcomings should be ascribed to a dearth of sufficiently varied methodologies and disciplinary backgrounds. The methodologies and backgrounds Runner, Levon and I bring to the table add to that intellectual diversity, strengthening the necessary climate of mutual critique and constructive feedback needed in a successful, empirical research programme.
2.3 The non-discrete model of Gay speech

Although I have argued in §2.1 after Zwicky that GS speech is not a discrete linguistic system, this claim is controversial. Eckert (2002) explicitly supports the usually tacit assumption that GS speech is a "target" of linguistic indexing – a formal system towards which at least some populations of men approximate their speech. This claim derives from Eckert's belief that labels such as Gay and Straight are salient in identity politics and serve as anchors for the clustering of various linguistic indexes into discrete wholes; but although I support this view of identity labels as anchors, I do not think it follows from the first half of this belief that the "clusters" must form discrete wholes. More importantly, the empirical facts do not support this claim.

Smyth & Rogers (2002) also view GS speech as aimed at a discrete target when they investigate the extent to which GS speakers sound more female than male, or approximate female forms rather than male forms. One instructive comment:

[F]ew sissy-sounding boys spend large amounts of time with gay or gay-sounding men. So boys who acquire these features must typically be getting them from their female family, friends, teachers and other role models. (§3)

Smyth & Rogers expend much ink arguing against a bimodal construction of gendered language and in favour of a continuum model; I, in contrast, see the continuum model as uncontroversial and instead question its unidimensionality. Yet even if we accept masculine and feminine as the poles of the gender continuum against which GS speech is measured and perceived, there is cause for concern with the characterisation in (1). Smyth & Rogers did not compare GS and female speakers' performance of several of the variables they measured, and therefore at best demonstrated that GS
speech is more feminine than SS speech, not that it is "like female speakers\(^9\)", which is a much stronger theoretical claim.

Assuming that the distribution of male and female gendered language is more or less bimodal, we may represent it pictorially in a model of multiple indexes (necessarily flattened into two dimensions) as in Fig. 1, where the grey circles are stereotypic targets of speech approximation around which actual data points are heavily clustered:

![Fig. 1: masculine/feminine speech](image)

Data points very near the grey circles are easily identified as masculine or feminine, while those in the middle are likely to be judged as gender nonconformist.

Now, consider that from the perspective of a data point within the M circle, other data points close to the M circle but a little to its right will still sound "feminine", even though they are nowhere near the F circle. We begin to see the danger of using the argument style of Smyth & Rogers

---

\(^9\) It should be noted that Smyth & Rogers (2002) report no significant difference in formant frequency between GS and SS speakers, despite the fact that gender differences in formant frequencies arise in the same age group as GS/SS differences. This indicates that the authors are aware that GS speech differs ontologically from gendered speech in at least one important way. My issue is thus with their terminology, not with their findings.
Fig. 2: Seattle's skyline as viewed from Kerry Park, on the south face of Queen Anne Hill

Stunning views of Seattle's skyline, with the Space Needle front and centre, may easily be captured from Kerry Park, 286 feet up the downtown-facing slope of the city's steepest and highest hill. This shot exemplifies the perspective problem which colours the perception of both hearers and researchers investigating gender and sexuality corollaries in speech. Columbia Seafirst Center, on the far left, is 932.4 feet tall, while the Space Needle, dwarfing every other building in the photo as well as in popular imagination, stands at only 605 feet. But Columbia Seafirst Center is three and a half miles' bike ride away from Kerry Park, down Queen Anne Hill, past the Denny Regrade, and
up the slope of First Hill, two full miles further away than the Space Needle. This succinctly
demonstrates an instantiation of the perception effect.

Let us return to Fig. 1, recasting it in terms of GS and SS speech:

Again, the grey circles are stereotypical areas that sound "totally Straight" or "totally Gay\textsuperscript{10}. But in
Fig. 3 the G circle does not serve as a target of indexing, contra Eckert. Instead, the dotted line
represents a threshold beyond which speakers may be judged GS, regardless of which index or
variable they employ, and irrespective of their possibly still great distance from the stereotype of the
G circle and still significant proximity to the S circle. This is analogous to a threshold of height
above which a building in the location of the Space Needle is perceived as a part of the rest of the
skyline, despite its distance from the skyline and proximity to the vantage point.

Why should GS speakers cluster around the S circle rather than the G circle? There are
many reasons. First, the G circle should be seen as a cluster of multiple GS indexes. But in line

\footnote{Since these stereotypes have different meanings for different individuals, there is necessary simplification here.}
with Zwicky, the *fact* of divergence from SS speech is important, not the number of indexes, since any one salient index marks a speaker as GS. Since there is little pressure to adopt multiple GS indexes, indeed quite to the contrary a heavy stigma on GS speech, an economy emerges whose points of equilibrium are close to the S circle, not to the G circle. Hearers are still adept at identifying GS speech, because they can identify many different salient features, only a subset of which marks each GS data point. Second, GS patterns in children at least have been shown by Smyth & Rogers (2002) not to parallel gender patterns entirely, indicating that GS boys at least in part index themselves according to masculine-sounding exemplars. Third, GS children are not exposed to GS adults as a general rule, and therefore cannot be seen as indexing themselves according to GS exemplars.

### 2.4 The empirical facts

The model in §2.3 predicts much in the distribution of my data. A non-discrete model (NDM) would predict less variation in the SS than in the GS task. I have already noted that this was true impressionistically. Quantitatively, the standard deviation of *all* dependent variables I analysed was lower in the SS than in the GS task. My NDM also predicts that there should be statistically significant generalisations that are true only for subpopulations. Many instances of this were detailed in §1.4-5. The NDM also predicts that participants might find the GS task more difficult than the SS task, because of the lack of an accessible, frequently deployed and well-learnt GS linguistic system. One participant, who I subjectively believe sounds quite Gay in everyday life, unelicitedly provided me a statement of precisely this type: he found the GS task very difficult, and
the SS task quite easy.

The NDM correctly predicts that features such as vowel lowering, laxing and rounding might apply only to some of the participants. The NDM is consistent both with previous findings that GS speech is "careful" and "articulate", and with my finding that it is quite the opposite: the two characterisations apply in different contexts to different subpopulations. I have taken a first step in extricating those confounding factors by identifying the factor of emotional rendition in my analysis.

The non-discrete GS data from my study, it may be argued, are to be expected in light of the fact that I asked each participant to do a normal, everyday rendition of GS, in contrast to Crist's (1997) asking for their "queeniest, most flaming gay stereotype". But Crist also failed to extract major, significant, discrete findings from his data, which is remarkable considering that his participants may well have been targetting a discrete target such as the G circle in Fig. 3. But the NDM correctly predicts Crist's data as well as mine.

The NDM is consistent both with previous failures to correlate average pitch with GS speech, and with a real such correlation. It would only be necessary to extricate the proper subpopulation, and here is where previous research has been inadequate. Further, the NDM serves as a strong warning against too quickly rejecting correlations on grounds of failure to "prove" them. If particular features are significant for only certain speaker subpopulations, and salient for only certain hearer subpopulations, then testing them quantitatively across the whole population and subsequently throwing them out is akin to throwing the baby out with the bathwater. Indeed, quantitative methods applied to language are often too constrained: since language cannot be
normalised or randomised, there is no reason not to apply quantitative measures to variables specifically picked out because they have been identified already by qualitative measures within the same data set. This is the reason for microanalysis and ethnography in the social sciences!

Methodologically, of course, the NDM is much more difficult to use in creating empirically testable hypotheses, because it requires the extrication and disentanglement of far more confounding variables than the discrete model does. Indeed, although I planned on a study including both a speaking and a listening task, as in previous studies, the sheer amount of detail and analysis prevented my getting the second part done within the scope of my human-subjects approval period!

But the discrete model has failed, and must be rejected. I believe that some of the methods and suggestions I have presented and cited in this paper will be useful in moving ahead with a rigorous research programme on GS speech within linguistics.

References


