

# 1 Introduction

A sizable body of research has been devoted to the investigation of the sociolinguistic perception of gay male speech. As early as the mid 1990s, reviews were being published on the potentially—and ostensibly—unique features of gay language (Jacobs, 1996). The results of these initial lines of inquiry, were, and remain, extremely inconclusive. Levon's paper, *Sexuality in context: Variation and the sociolinguistic perception of identity* (2007) represents a furthering of the aim of understanding what phonetic cues listeners draw on in the evaluation of a speaker's sexuality. In what follows I will first briefly outline research that had been done prior to that paper, and, second, I will explore what made Levon's contribution particularly unique, in order to provide background for the current study, which is a replication of that 2007 experiment using original materials within a novel experimental setting and subject pool, Amazon's Mechanical Turk platform.

## 1.1 Work prior to 2007

It is common for native speakers of American English to have some kind of intuition that it is possible to determine the sexuality of a speaker by listening to their language use. It is perhaps understandable, then, that researchers working within the diverse fields of sociology, linguistic anthropology, and speech pathology have attempted to answer the question: what linguistic cues are present in the speech stream of gay men's language that allow for such a determination to be made? After all, Gaudio (1994) found that the 13 listeners in his perception experiment were able to correctly determine the sexuality of seven out of eight of the male speakers that contributed speech stimuli. Similarly, (Linville, 1998) found that the listeners in her study were able to correctly identify the sexuality of a speaker's sexuality 79.6% of the time.

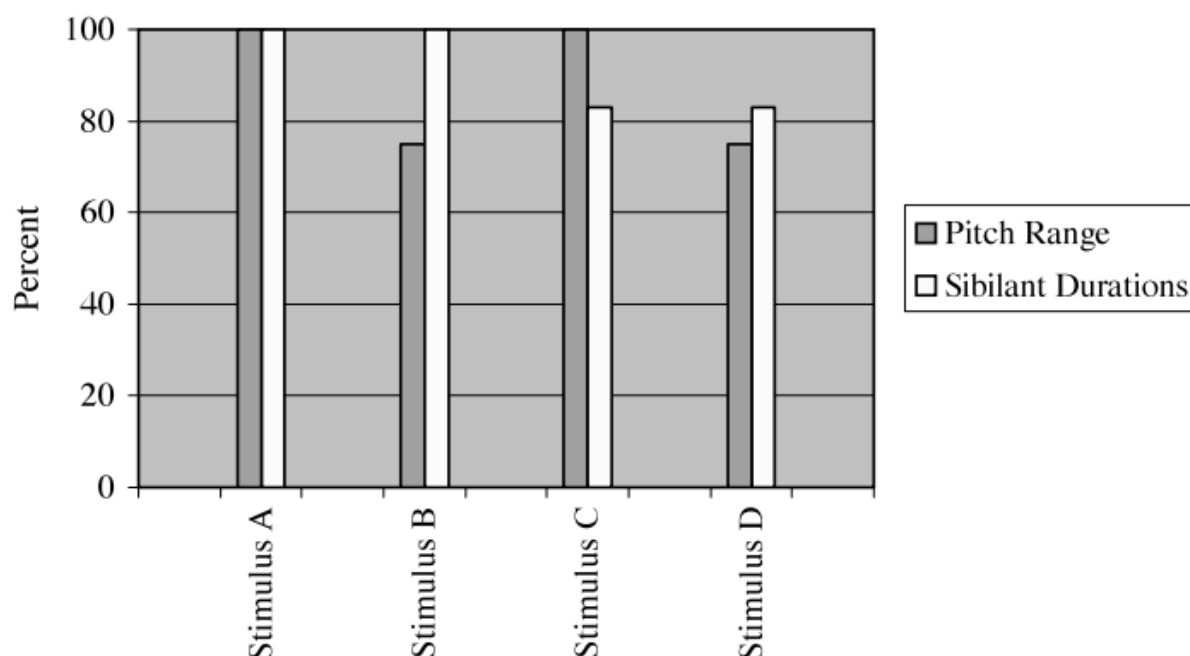
Because Levon's study focuses on two phonetic characteristics (what Levon calls features) in particular, PITCH RANGE and SIBILANT DURATION, I will frame the present discussion around those phonetic characteristics. Both variables were thought to play a role in the evaluation of a speaker's sexuality, perhaps due to folk ideological assumptions about the speech of gay men's speech. Gaudio, for example, found no significant differences in the pitch ranges of the language of self-identified gay and straight speakers. Similarly, Smyth, Jacobs, and Rogers (2003) found no significant relationship between either mean F0 or pitch range and the evaluation of a speaker's sexuality by a listening subject population. There has existed—and still exists—a popular understanding of gay men's speech as being “lispy”. In order to investigate this potential association, Jacobs, Smyth, and Rogers (2000) conducted an experiment in which listeners were asked to rate how gay the speech they were asked to listen to sounded. They found that the recordings that were rated as more “gay-sounding” had significantly longer duration of both /s/ and /z/. Related to this finding is Linville's 1998 finding that attribution of gayness to a speaker's voice is significantly related to the spectral characteristics of the voiceless alveolar fricative, /s/, namely the length of frication and higher peak frequency.

## 2 Levon 2007

In his discussion of previous work, Levon devotes some time to outlining what he sees as two problems with the previously predominate methodologies. First, the four studies I have summarized above essentially were run as follows: listeners were shown to be somewhat adept at identifying the sexuality of the speakers they were asked to listen to; the researchers then looked at all self-identified "gay speech", in order to see what were its unique phonetic properties. As Levon writes: "the methodology employed has implicitly assumed gay-identified speech to be an essential component of gay people, and then set out to quantify it" (536). Second, he argues that it is impossible to determine in any substantive way what cues listeners are attuned to in the task of identifying a speaker's sexuality; there are, Levon contends, "potentially hundreds of linguistic differences among the samples" (536). A more nuanced approach is necessary, he contends—one that aims to determine the role that specific cues play in the task at hand, as well as their interactions with one another. By using novel acoustic manipulation techniques, Levon aims to create an environment in which variation is controlled for.

### 2.0.1 Stimuli and experimental design

Recordings of four speakers reading the same passage, a story of a typical occurrence on a crowded subway platform in Manhattan, were created. Three of these speakers were rated by a group of ten graduate students in Linguistics at New York University as sounding "straight" and "masculine". One recording was made by a speaker that was rated by the same group of graduate students as sounding "extremely gay" and "extremely effeminate". Crucially, the speech of one of the "straight-sounding" speakers as well as the "gay-sounding" speaker were acoustically manipulated. The speech of the speaker that was rated as sounding gay was altered such that three stimuli in addition to the original recording were produced, stimuli A, B, C, and D. Stimulus A consisted of the unaltered recording. Stimulus B consisted of speech in which the pitch range was narrowed by 25%, the magnitude of potential individual difference in pitch range between speakers as reported in Levon (2004). Stimulus C consisted of speech in which the duration of the sibilants /s/, /z/, and /ʃ/ were shortened by 17%, the threshold at which Jacobs et al. (2000) and Linville (1998) reported a change in listeners' perception of a speaker's sexuality. Stimulus D consisted of speech in which both pitch range was narrowed and the duration of the aforementioned sibilants was shortened. Figure 1 shows the "gay-derived" stimuli A-D.



A: [+wide] pitch range, [+long] sibilant duration (*original*)

B: [−wide] pitch range, [+long] sibilant duration

C: [+wide] pitch range, [−long] sibilant duration

D: [−wide] pitch range, [−long] sibilant duration

**Figure 1:** A reproduction of Levon’s Figure 1, showing the gay-derived stimuli A-D

Four additional stimuli were created from one of the recordings that was rated as belonging to a “straight-sounding” speaker, similar but in an opposite fashion as the “gay-derived” stimuli. Stimulus E consisted of the unaltered recording. Stimulus F consisted of speech in which the pitch range was widened by 25%. Stimulus G consisted of speech in which sibilants were lengthened by 17%. Stimulus H consisted of speech in which both pitch range was widened and sibilants were lengthened.

The recordings were manipulated in this way in order to determine the relative effect of the relevant phonetic difference on the evaluation of the speakers’ sexuality. By presenting a subject pool with both “gay-derived” and “straight-derived” stimuli, which were manipulated in an opposite fashion, the relative significance of each feature, in isolation and in concert with the other, may be able to be determined, particularly in relation to the character of the original voice, i.e. whether it was already perceived as sounding gay or straight.

123 undergraduate students at two New York City universities were recruited to participate in the

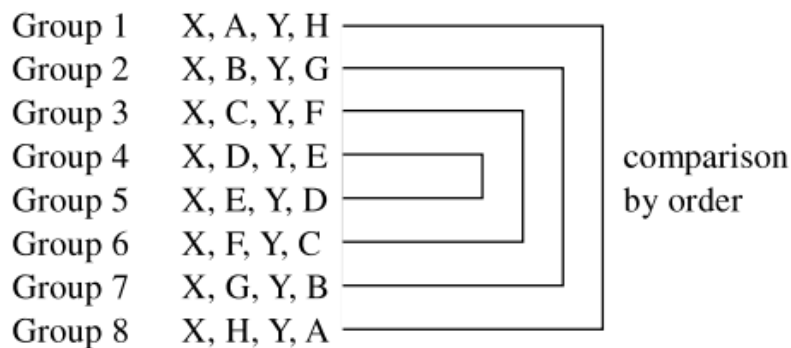
original study. This listener population was split into eight groups. Each of these groups heard four recordings and were asked to offer their judgments about qualities of the speakers they heard on a ten-dimensional, seven-point Likert scale, reproduced below in Figure 2.

			1 = extremely			4 = neutral		7 = extremely		
1	(S)	generous	1	2	3	4	5	6	7	greedy
2	(S)	lazy	1	2	3	4	5	6	7	hardworking
3		prudish	1	2	3	4	5	6	7	promiscuous
4	(G)	effeminate	1	2	3	4	5	6	7	masculine
5	(S)	aloof	1	2	3	4	5	6	7	friendly
6	(G)	straight	1	2	3	4	5	6	7	gay
7		neat	1	2	3	4	5	6	7	messy
8	(S)	savvy	1	2	3	4	5	6	7	naïve
9	(S)	kind	1	2	3	4	5	6	7	mean
10	(G)	genuine	1	2	3	4	5	6	7	fake

(S) – adapted from Scherer (1972); (G) adapted from Gaudio (1994)

**Figure 2:** A reproduction of Levon's Table 1, showing the ten dimensions on which listeners were asked to judge each speaker, based on their voice

The first and third recordings that all eight groups heard were the same, labeled X and Y in Levon's original study; both of these recordings were judged by the same initial listening pool as sounding both straight and masculine. The second and fourth recordings were varied, however; the ordering of the recordings played for each group is shown in Figure 3.



**Figure 3:** A reproduction of Levon's Table 2, showing the ordering of the recordings that listeners in groups one through eight were asked to respond to

Note that groups one through four heard different recordings, while groups five through eight heard these same recordings in the opposite order. This does mean that not every possible combination

of stimulus was encountered by any one person; this is because it would be impossible to present a listener with two recordings of the same speaker, as they would immediately recognize that it was the same person. Because of this impossibility, ratings were compared across the four groups that heard different recordings.

### 2.0.2 Original results

One of the key results of Levon's study was that a significant main effect of stimulus order was found for the "gay-derived" recordings. That is, if a listener heard the "gay-derived" recording (A-D) fourth, they rated it, on average, as sounding more gay than if the listener heard it second. Levon also conducted a pairwise comparison to determine what effect, if any, the acoustic manipulation had on listener ratings on the gay-straight and effeminate-masculine scales. He found that only when both sibilants were shortened and pitch range was narrowed did listeners rate the speaker as sounding significantly more masculine. As far as the gay-straight scale is concerned, he found that the key alteration was the pitch range manipulation; the shortening of the sibilants seemed to play no role in listeners' evaluation of the speaker's sexuality. Critically, Levon found no stimulus or order effect for the "straight-derived" speech. He does not report on any of the other eight dimensions on which listeners offered their evaluations of the recordings.

## 3 Replication

### 3.1 Introduction

A replication of Levon's original study was carried out under novel circumstances. 120 subjects were recruited on Amazon's Mechanical Turk platform. A web-based implementation of Levon's experiment design was created using the original stimuli, including decoy recordings X and Y. Exclusion criteria were established by use of an audio and attention check, in which listeners were forced to listen to a short interaction about the quality of a salad and were prompted to respond with what the interaction was about; five participants were excluded from analysis for having given more than one incorrect answer, perhaps indicating that they were a non-human actor or that their audio was not functioning properly. As in the original experiment, participants were divided into eight listening groups and were presented with the corresponding stimuli in the appropriate order. Participants were unable to rate the speech without having listened to it in its entirety.

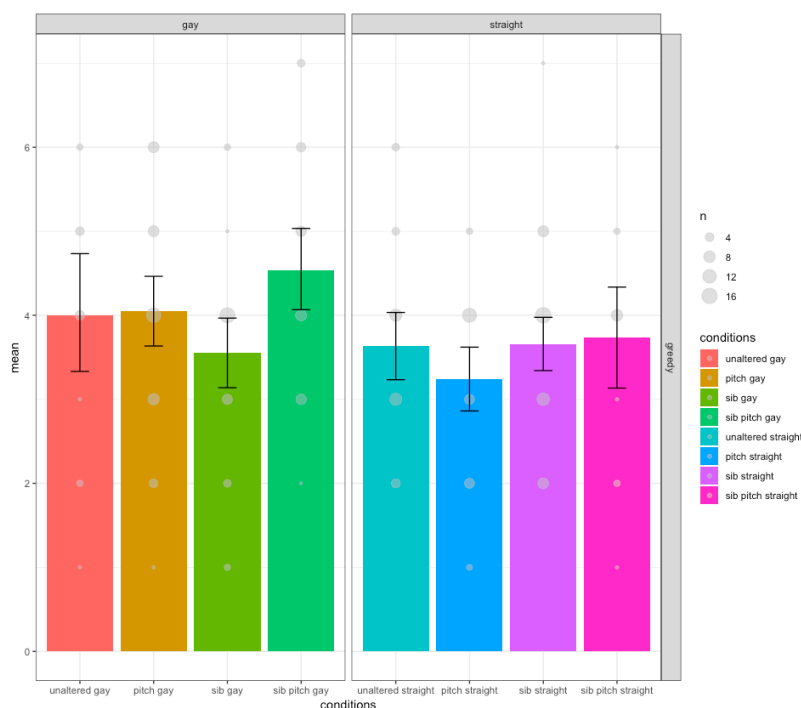
### 3.2 Results

In order to determine the effect of order of presentation and stimulus condition, mixed-effects ordinal regression models were conducted. Each predicted the 1-7 rating value for each adjective pair from a three-way interaction of order of stimulus, condition of the original recording (i.e. whether the recording was "gay-derived" or "straight-derived"), as well as a factor of number of features

that were altered. By-participant random slopes were computed. The results of these models are as follows: for no adjective pair was any interaction of predictors significant. Additionally, for no adjective pair, including straight-gay and effeminate-masculine, did the order of the stimulus presentation reach significance. However, a main effect of original condition was found for four adjective pairs: generous-greedy, effeminate-masculine, straight-gay, and kind-mean. Effects for each of these pairs is provided below.

### 3.2.1 Generous-greedy

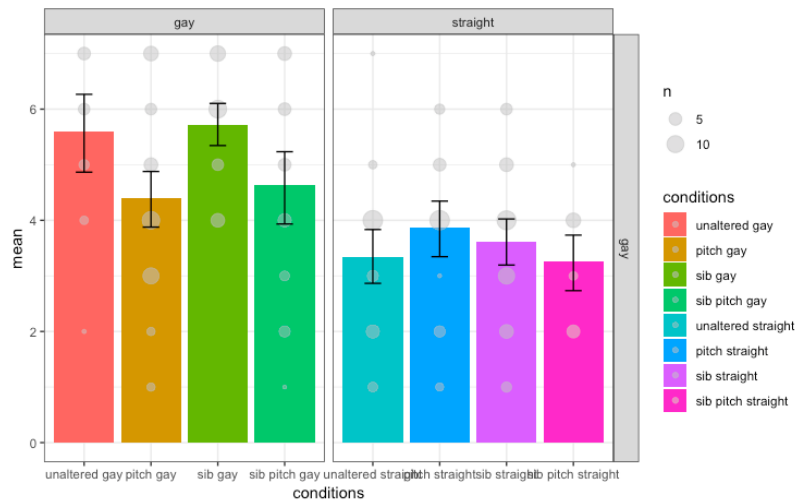
An effect of original stimulus condition for this adjective pair was found ( $\beta = -1.44720$ ,  $SE = 0.58$ ,  $p < 0.05$ ). That is, listeners were more likely to rate the “gay-derived” speech as sounding more “greedy”. A plot of listener ratings for each stimulus, irrespective of stimulus order, is provided in Figure 4.



**Figure 4:** Listener ratings for the generous-greedy adjective pair, where a higher rating is closer to the “greedy” pole

### 3.2.2 Straight-gay

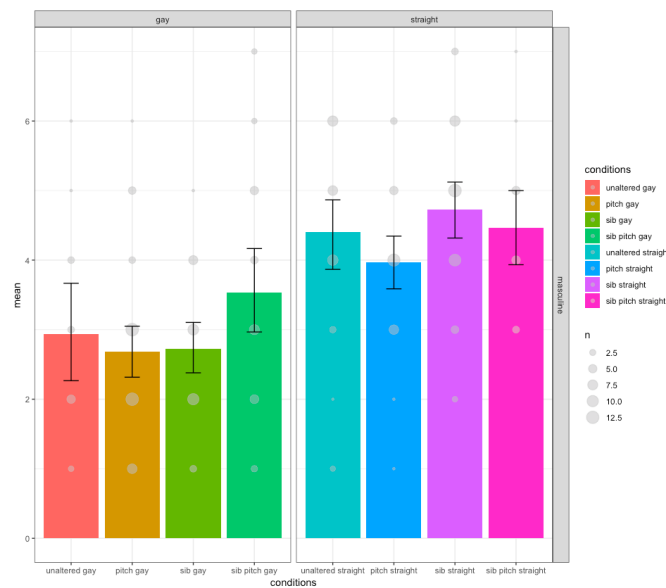
An effect of original stimulus condition for this adjective pair was found ( $\beta = -1.45838$ ,  $SE = 0.45296$ ,  $p < 0.01$ ). That is, listeners were more likely to rate the recordings in the predicted way: “gay-derived” speech was rated as sounding more “gay”. As in the previous section, a plot of listener ratings for each stimulus is provided below.



**Figure 5:** Listener ratings for the straight-gay adjective pair, where a higher rating is closer to the “gay” pole

### 3.2.3 Effeminate-masculine

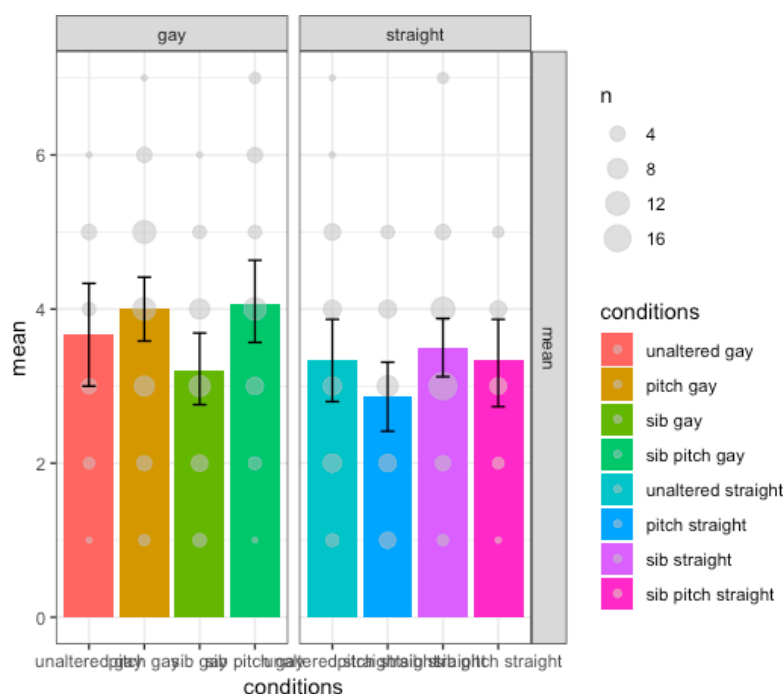
An effect of original stimulus condition for this adjective pair was found ( $\beta = 2.04628$ ,  $SE = 0.43111$ ,  $p < 0.001$ ). Listeners rated the “straight-derived” speech in the predicted way. A plot of listener ratings for each stimulus is provided below.



**Figure 6:** Listener ratings for the effeminate-masculine adjective pair, where a higher rating is closer to the “masculine” pole

### 3.2.4 Kind-mean

An effect of original stimulus condition for this adjective pair was found ( $\beta = -1.40174$ ,  $SE = 0.60112$ ,  $p < 0.05$ ). Listeners rated the “gay-derived” speech as sounding more “mean”. A plot of listener ratings for each stimulus is shown below.



**Figure 7:** Listener ratings for the kind-mean adjective pair, where a higher rating is closer to the “mean” pole

### 3.3 Discussion and future directions

Unlike Levon’s original results, neither a significant effect of stimulus order nor of number of features acoustically altered were found for either the “gay-derived” or “straight-derived” stimuli. It appears that perhaps there is something else about the voices apart from sibilant duration and/or pitch range that are affecting listener ratings in the predicted direction. It could be that other phonetic features are acting as clues. Unclear, then, is how one might go about disentangling the co-occurring phonetic features of each voice from the role that sibilant duration and pitch range play as cues to a speaker’s sexuality. One potentially illuminating extension of this research programme would be to carry out the experiment with a more robust sample of different voices. In this way, one would be able to control for effects of what I am calling the “stylistic totality” of the voice in question. That is, there are innumerable phonetic cues that occur in any given voice, and they occur in concert with every other cue—by carrying out the experiment with greater number of voices, one might be able to say with more certainty whether there was something specific about a particular voice that influenced listener ratings or whether a more general phenomenon was being observed.



Additionally, since listeners would encounter more than two critical stimuli trials, a more robust effect of stimulus order may be found.

## **4 Conclusion**

Levon's original experiment design continues to represent a novel contribution to the investigation of stereotype and the perception of sociolinguistic variation. A replication was carried out, which found no effect of acoustic manipulation on listener ratings of sexuality or gender, but only a main effect of original stimulus condition on the ratings for four adjective pairs, effeminate-masculine, straight-gay, kind-mean, and generous-greedy. Without some kind of post-hoc analysis, it is unclear what it is about the individual voices that resulted in this significant difference in listener ratings. A similar experiment could be carried out in which more voices were included; in this way the effect of the "stylistic totality" of any given voice could be taken into account and more robust and generalizable findings about the nature of sexuality evaluation based on speaker voice may be arrived at.

## References

- Gaudio, R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American speech*, 69(1), 30–57.
- Jacobs, G. (1996). Lesbian and gay male language use: A critical review of the literature. *American speech*, 71(1), 49–71.
- Jacobs, G., Smyth, R., & Rogers, H. (2000). Language and sexuality: Searching for the phonetic correlates of gay-and straight-sounding male voices. *Toronto Working Papers in Linguistics*, 18.
- Levon, E. (2004). Examining a gay prosody: Issues in theory, methodology and identity. *Lavender Languages and Linguistics XI, American University, Washington, DC*.
- Levon, E. (2007). Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society*, 36(4), 533–554.
- Linville, S. E. (1998). Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Folia phoniatrica et logopaedica*, 50(1), 35–48.
- Smyth, R., Jacobs, G., & Rogers, H. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society*, 32(3), 329–350.