Duration of Onset Consonants in Gay Male Stereotyped Speech

Sean Crist

Gaudio (1994) found that listeners are largely accurate in identifying the sexual orientation of male speakers from recordings of a read passage not marked for gay content. In search of the phonetic cues which listeners use in this identification, Gaudio studied F0 differences between gay and straight male speakers. Gaudio subjected the recorded speech to multiple kinds of F0 analysis (mean F0, median F0, etc., as well as several different measures of pitch dynamism), and concluded that range and variability of F0 do not appear to correlate with sexual orientation, at least by any of the measures used.

If listeners can identify the sexual orientation of speakers with reasonable reliability, then what are the salient phonetic cues, if not range and variability of F0? One possibility is that F0 plays an important role in sexual orientation identification, but does so in a way which has not yet been identified. Another possibility is that some or all of the salient cues are unrelated to F0. Without ruling out the former possibility, the present study[1] takes a preliminary look at some variables not related to F0. I will argue that the performance of a gay male stereotyped voice involves the systematic lengthening of certain consonants.

My general methodology differs from Gaudio's in one major regard. Gaudio recorded gay and straight men speaking in their ordinary voices, and then compared the two groups. By contrast, the speakers in the present study (who included both gay and straight men) were asked to read the same passage twice, once in an "ordinary" voice, and again in a "queeny" voice.

One must be extremely cautious in making inferences about natural speech on the basis of a consciously performed stereotype: there is no guarantee that the cues used by listeners to identify gay male speakers are the same cues which speakers employ to deliberately produce an exaggerated gay male stereotype. In spite of this shortcoming, there were two reasons for my choosing this methodology over one involving the comparison of the ordinary read speech of gay and straight male speakers. First, the question of how speakers perform a gay male stereotype is an interesting one in its own right; and when our knowledge reaches a state where it is possible to do so, it will be of interest to compare the cues used in stereotype production against

---

[1]This paper has benefited from comments from a long list of individuals. Special note should be given to Gene Buckley, Rolf Noyer, Mark Liberman, William Labov, and two anonymous reviewers. Any errors in this work are of course my own.

the actual cues used in sexual orientation identification of males. Second, and more importantly, we presently simply do not know what variables are involved in either of these two tasks (stereotype production vs. identification of sexual orientation on the basis of natural speech). It is a reasonable guess that the variables involved in stereotyped speech are likely to undergo a more extreme exaggeration than those in the former, thus lending themselves to easier identification by the linguist. Once some of the variables involved are identified, it becomes a much easier task to determine whether a more subtle variation along the same dimensions helps listeners to distinguish gay men from straight men in non-stereotyped speech. The present study is thus an instance of searching where the light is best, which is not an unreasonable place to look when next to nothing is known.

Since there are many factors which can affect segment duration, this study restricts its scope to the onsets of word-initial syllables bearing primary stress, thus controlling for syllable position and stress. The possibility that duration of consonants in other environments plays a role in the performance of a gay male stereotype is not ruled out and is left for future study.

The present study involves two separate experiments.

1. First Experiment

The subjects of the first experiment were three white males who were native speakers of American English. One self-identified as gay and two self-identified as straight; all had at least a bachelor's degree[2]. Both of the straight speakers had some graduate-level training as linguists, but the gay speaker had none.

The task given to the subjects was to read a short printed story, first in "your ordinary voice", and then in "the queeniest, most flaming gay stereotype you can do." The story was written to exclude any gay-specific content, and was intended to resemble extemporaneous speech, while

[2]Actually, a total of seven subjects were recorded for the first experiment. Three of the subjects were more or less arbitrarily excluded from analysis because of the time required to segment the passages (approximately three hours per recording, for a total of around six hours per speaker). The speech of four subjects was segmented, but the data of one speaker were excluded from the analysis since they were wildly unlike those of the other speakers in nearly every regard, even for variables where the other speakers were consistent both across speakers and across stereotyped and non-stereotyped speech. The recordings of this subject were played for a group of a half dozen linguists, who agreed with my subjective assessment that the speaker had some mild sort of speech disorder.

including as many tokens of the types under study as reasonably possible[3].

Subjects were digitally recorded at a sampling rate of 8000 Hz on a Sun Workstation. Measurements in milliseconds of certain segments were later made from spectrograms of the recordings using the xwaves software package. The following segments were measured when they occurred in the onsets of stressed word-initial syllables:

- •Closure and aspiration time for /p/, /t/, /k/ in the environment #_V (= 6 token types)
- •Frication time for /s/ in the environments #_V and #_mV (= 2 token types)
- •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 (= 21 token types)
- •Frication time for /f/ in the environments #_V, #_lV, #_rV (= 3 token types)
- •Duration of /h/ and /l/ in the environment #_V (= 2 token types)

The rationale for this choice of token types was as follows. In preliminary recordings, there seemed to be differences in the duration /l/ (the effect was particularly pronounced in the word 'believe', but this word was later excluded from measurement when the experiment was restricted to word-initial onsets). Further, an similar effect was informally noticed for word-initial /s/ as it occurs in complex clusters. There is a common stereotype

---

[3] The text of the story used in the first experiment was as follows: "You wouldn't believe what just happened! I was just sitting here studying, and it was getting pretty late, and I was going to go to bed here pretty soon. But then I started hearing these people screaming out in the street. So I got up, and I was going to yell out the window, 'Will you please hold it down out there!' But as soon as I poked my head out, I smelled smoke, and you know that ski store down at the corner? It was all full of flames. There were all these people in the apartments upstairs screaming out of the windows; they must have been trapped. I was scared that the fire might spread down the street to my place too. Then I heard sirens screaming, and all these cop cars and fire trucks pulled up. The firemen went up on ladders and helped all the people get out. One girl looked like she had bad burns on her skin, and this other guy fell, and the ambulance guys had to put a splint on his leg. I could see the guys down on the ground; they were having some kind of problem with the fire hydrant, but they finally got the hoses hooked up to the spouts, and then they went up and poked a hole in the roof with a big metal kind of stick, and they sprayed tons and tons of water in. It took them better than two hours to get the fire out. You know that Spanish student down the hall from me? Later, he told me he heard the owner set the fire himself. The whole thing was a big scam to get the insurance money. Unbelievable!"

that gay male speech is ‘breathy’ or ‘lisping’, which could conceivably be realized in the durations of aspiration and of /s/, respectively (i.e., instead of or in addition to actual laryngeal breathiness on vowels or particular spectral qualities of /s/); a group of onsets was chosen to study the duration of aspiration and of /s/ in detail. Another fricative, /f/, was chosen to determine whether any effects discovered with /s/ were obviously true of other fricatives as well. Naturally, this selection omits a large group of possible English onsets. The overall intent was to cast as wide a net as possible within the bounds of a single, manageable study.

Excluded from measurement were tokens occurring in an environment with a phonetically confounding factor which made measurement impossible or of questionable salience. For example, /s/-frication time was not measured in the case where the word-initial /s/ is preceded by a word-final /s/, as in “…upstairs screaming”, since there is of course no basis for dividing the duration of the single observed region of frication between the two /s/ segments (the story script was written with the general avoidance of such cases in mind, but there remained some problematic cases which were explicitly excluded from measurement).

Altogether, a total of 41 token types were measured[4]. The script contains 176 tokens for which values were sought. Instances where measurements could not be obtained (because of speaker misreadings of the script, unresolvable ambiguity in the spectrogram, etc.) were not coded and were excluded from analysis.

1.1. Analysis of the First Experiment

Boxplots were produced for each of the 37 duration token types using DataDesk, a Macintosh statistical package. The stereotyped vs. non-stereotyped recordings for each speaker were compared; and these three pairs of boxplots were in turn compared, giving a total of six boxplots per variable. The boxplots were examined in search of cases where the gay

---

[4] Additionally, the four following token types were coded with binary values for application or non-application of the relevant rule:

- •Flapping: t, d → □ / ʌ□_ʌ
- •gonna-contraction (going to → gonna)
- •Final t deletion in the word “just”
- •Use of /-in/ rather than /-□/ for the -ing suffix

However, it became immediately apparent that there were too few of these token types to make any kind of analysis possible; a substantially larger corpus would be needed for the investigation of these variables.

stereotyped measurements varied from the non-stereotyped measurements in a consistent way across speakers.

For the great majority of variables, no regularity of this sort was observed, although it should be noted that some patterns may have been obscured by the relatively small number of tokens in each category. However, there were a few important exceptions where such an apparent pattern was to be seen, i.e. variables which might serve as cues for the sexual orientation of the speaker.

First, duration of /l/ in the environment #_V was consistently longer across speakers in stereotyped than in non-stereotyped speech.

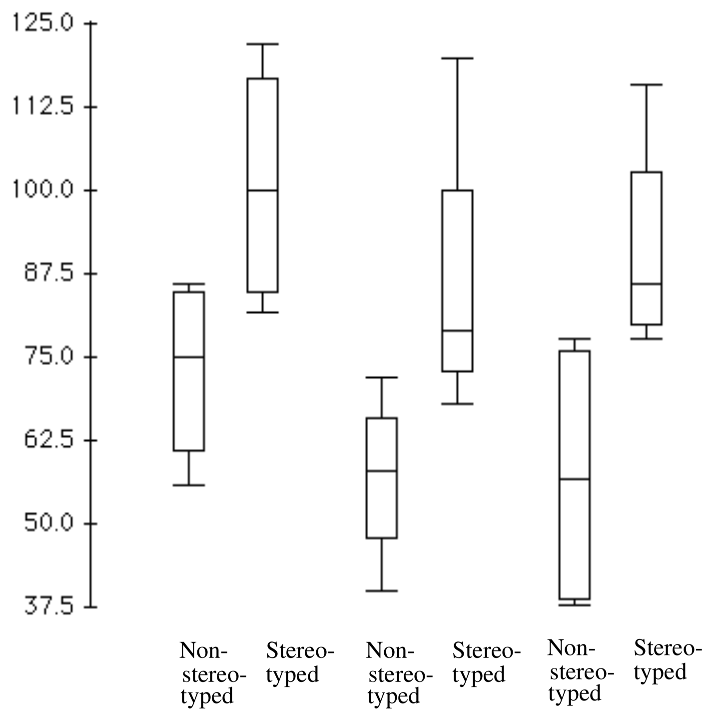


Figure 1: Duration of /#l/ (n=5 per boxplot)

Using the DataDesk statistical analysis package for the Macintosh, pooled t-tests were performed for each speaker comparing $\mu1$-$\mu2$, with a Ha of $\mu1 \neq \mu2$, for the stereotyped and non-stereotyped /l/ durations. For none of the individual speakers was the result significant at p=.05 (p=.0579, .0949,

.0949, for the three speakers respectively), which is not surprising given the relatively small number of tokens. However, when the non-stereotyped tokens for the three speakers are pooled (giving a total of 15 tokens in each of the two categories) the difference is significant (p=.0015)[5].

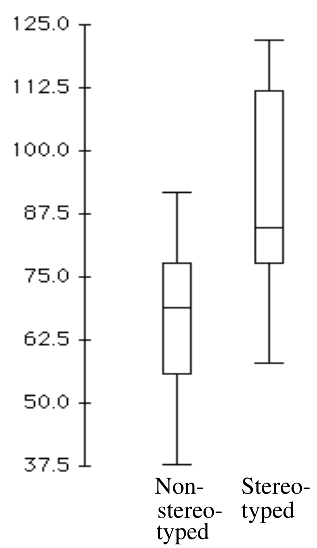
Second, frication of /s/ in #skV clusters was consistently longer in

---

[5] Since the means for stereotyped and non-stereotyped /l/ duration differ across speakers, it could be argued that some kind of normalization should be applied before lumping speakers together in this way. As an aside, it is not obvious what kind of normalization should be used in the case where each speaker has *two* means, which must somehow be considered in tandem. In any case, I think the fact that the lumped data differ significantly (in spite of not being normalized) actually argues that the difference is a robust one.

Suppose that speakers do in fact systematically lengthen /l/ when producing a gay male stereotype, but that speakers differ with what durations they count as a "long" or "short" /l/ (such a difference between speakers could be as simple as a function of overall speaking rate). If this is the case, then we should expect that a failure to normalize the data would tend to *obscure* this pattern when speakers are lumped, rather than create a false pattern where none actually exists. If the pattern remains in spite of failure to normalize, it would seem to mean that the pattern is robust.

When the raw data for /l/ are lumped, the resulting boxplot is as follows:



Durations of /#lV/for all three speakers (p=.0015) stereotyped than in non-stereotyped speech.

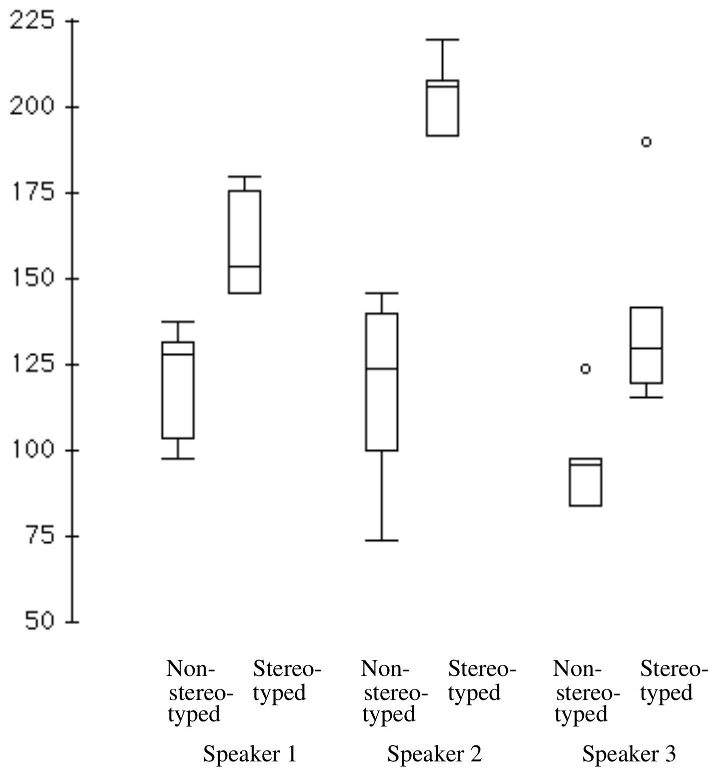


Figure 2: Duration of /#sk/ frication[6]. (n=5 per boxplot)

Again, t-tests were performed comparing each speaker's stereotyped and non-stereotyped speech. For each individual, the difference was significant (p=.0058, .0003, .0239, respectively).

---

[6]The DataDesk software package uses the ° symbol in boxplots to mean "an extreme data point," i.e. one which falls outside the whiskers. The upper whisker extends to the highest data point not outside:

high hinge + 1.5 (high hinge - low hinge)

The lower whisker extends to the lowest data point not outside

low hinge - 1.5 (high hinge - low hinge)

The * symbol found in later boxplots represents "a very extreme data point"; the formulae are the same as those for "extreme data points", except that the coefficients are 3.0 rather than 1.5 (William Labov, personal communication).

The pattern found for #skV frication time might also hold true for #spV frication time; however, since there are only two tokens of #spV in the script, nothing conclusive can be said about this variable at present. The lumped #spV tokens for the three speakers produce a pattern at least consistent with that of #skV:

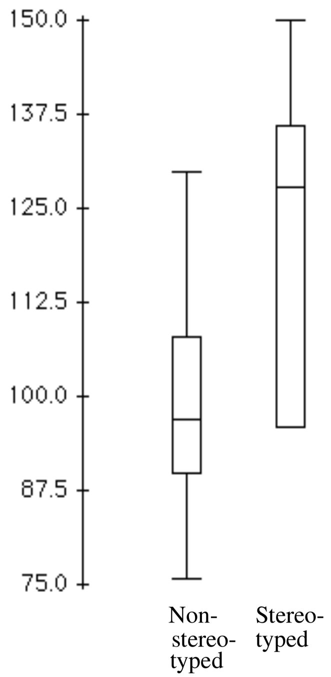


Figure 3: Lumped durations for /#sp/ frication

A t-test over these two groups is suggestive (p=.0803). This variable is studied in greater detail in the second experiment.

By contrast, frication time for #stV appears not to participate in the pattern of #skV (and possibly #spV):

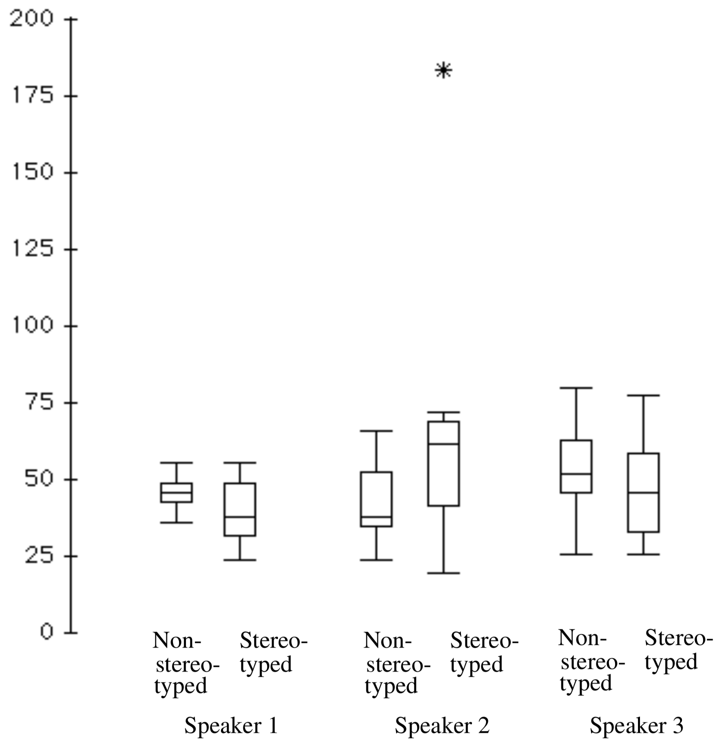


Figure 4: Duration of /#st/ frication.

2. Second Experiment

Once an apparent pattern with /l/, /sp/, and /sk/ (but not /st/) had been identified, a second experiment was constructed to obtain more tokens of these types to overcome the paucity of tokens in the first experiment. A new script was written[7] including 15 tokens of /l/, 6 of /sp/, 8 of /st/, and 4 of

[7] The text of the story used in the second experiment was as follows: "You wouldn't believe what just happened! I was just here studying, and I heard this big loud noise outside. So I was going to go yell out the window, 'Listen, will you please hold it down out there!' But then I looked out, and you know that ski store downstairs next door to me? Oh, my God, it was on fire! I guess the stairs were blocked, because the tenants upstairs looked like they were stuck up there. They were screaming out of the windows in Spanish, but I don't speak Spanish, so I

/sk/, all in the appropriate environment.

The subjects for the second experiment were three gay males. One was an undergraduate majoring in a field unrelated to linguistics (Speaker 4); a second was a professor in a field unrelated to linguistics, who was in his mid-thirties (Speaker 5); and the third was an elementary school counselor in his fifties (Speaker 6). All were Caucasian and were native speakers of American English. None had participated in the first experiment.

The methodology for the second experiment was identical to that of the first, with the exception that two of the speakers were recorded at a sampling rate of 16,000 Hz rather than 8000 Hz, in anticipation of a future study of the spectra of /s/.

2.1. Analysis of the Second Experiment

During segmentation, it was noticed that the durations of /l/ in the second script's three instances of *looks… like* were consistently very short. These tokens were separated out and analyzed separately. The data for /l/ are as follows shown in figure 5.

The data for Speakers 4 and 5 in figure 5 are consistent with those found in the first experiment, and the differences are significant for both speakers without any lumping of speakers. The p-value for Speaker 6 is unexpected, however; as will be seen below, this speaker is inconsistent with the other speakers for other variables.

---

didn't know what they were saying. And then I saw this one guy leap out of the window, and it looked to me like he really hurt his leg bad. It was really awful! And you know, there's not really a lot of space between that building and mine, so I was getting scared for myself, too. Anyway, then I saw a lot of red lights flashing, and finally the firemen got there from the station. They went up on ladders and got the people out, and also, they used that round thing that looks like a trampoline with a big red spot in the middle. And they had problems; I think the hoses were tangled up on those big spools they have on the trucks. But finally they got the hoses hooked up to the spouts on the fire hydrants. Then they went up and poked a hole in the roof with a big metal kind of stick, and it made the sparks fly all over the place. You know that Spanish student down the hall from me? Later, he told me he heard the owner of the ski store set the fire himself! The whole thing was a big scam to get the insurance money. Unbelievable!"

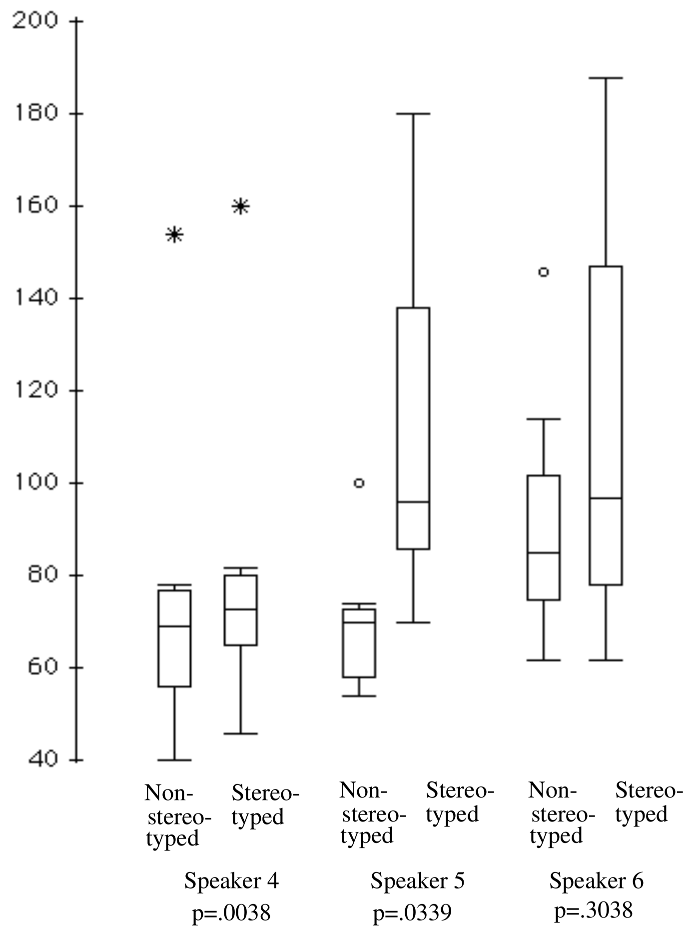


Figure 5: Duration of /#l/, excluding tokens in *looks… like.*

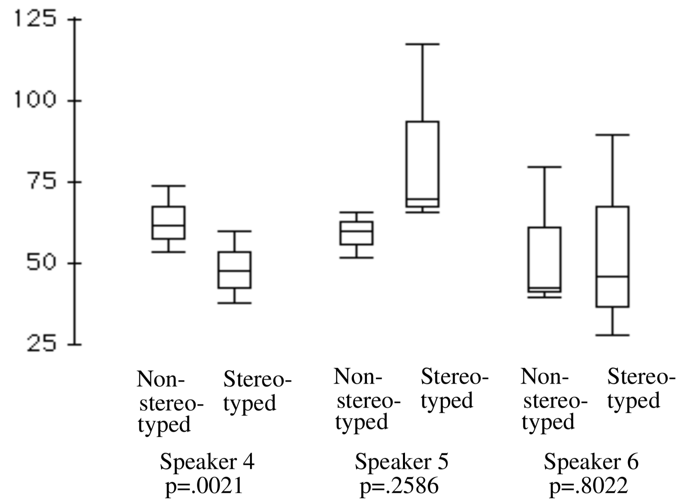


Figure 6: Duration of /#l/ in *looks… like.*

Figure 6 contains the durations for /l/ in *looks…like*. The pattern in Figures 1 and 5 does not seem to be not present here; the p-values are insignificant, except for Speaker 4, for whom the data are significant in the wrong direction. It appears that *look… like* does not participate in the gay-stereotyped lengthening of /l/, but what specific property of these words is responsible for this difference is not known. Perhaps these words are exempt from l-lengthening because of their prosodic status, or their status as quasi-function words, or perhaps there is something about the particular phonetic environment of /l/ in these words. The data at hand do not allow us to distinguish these hypotheses.

The data for /#sp/ are as follows:

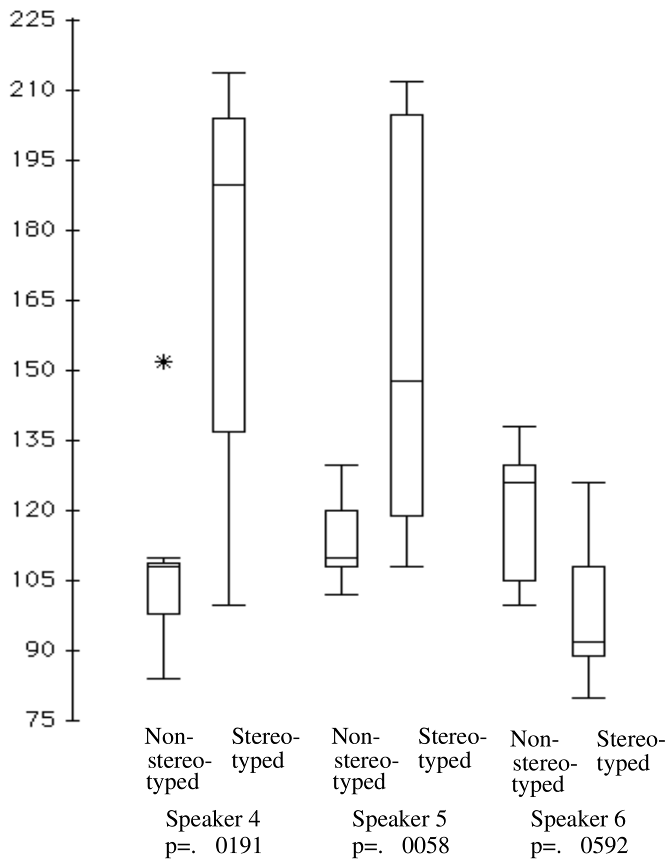


Figure 7: Durations for /#sp/

The data for Speakers 4 and 5 are consistent with the hypothesis that the gay male stereotype involves lengthening of /s/ in /#sp/, and are strongly significant without any lumping of speakers. The data for Speaker 6, however, runs counter to the hypothesis—nearly significantly so (the possibility that this anomaly was the result of a simple incorrect formatting of data was carefully disconfirmed).

The speakers' patterns for /#sp/ appear to hold true for /#sk/ as well:

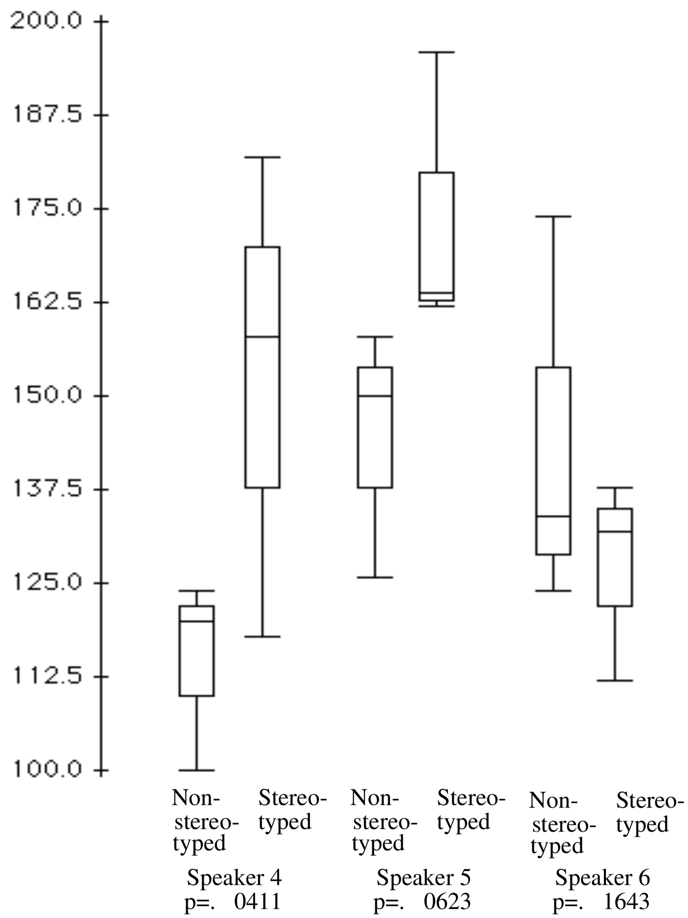


Figure 8: Durations for /#sk/

The data for Speakers 4 and 5 are consistent with those for Speakers 1, 2, and 3, as well as with those for /#sp/. The t-test for Speaker 5 revealed that his difference between stereotyped and non-stereotyped speech was not quite significant for this variable; this is not particularly distressing, however, since only 4 tokens of /#sk/ were elicited in the second experiment. The first experiment had found a strong enough pattern for /#sk/ that the second script was directed more at obtaining more tokens of /l/ and of /sp/.

As with /#sp/, the data for Speaker 6 are nearly the reverse of what would be expected.

The data for /#st/ are as follows:

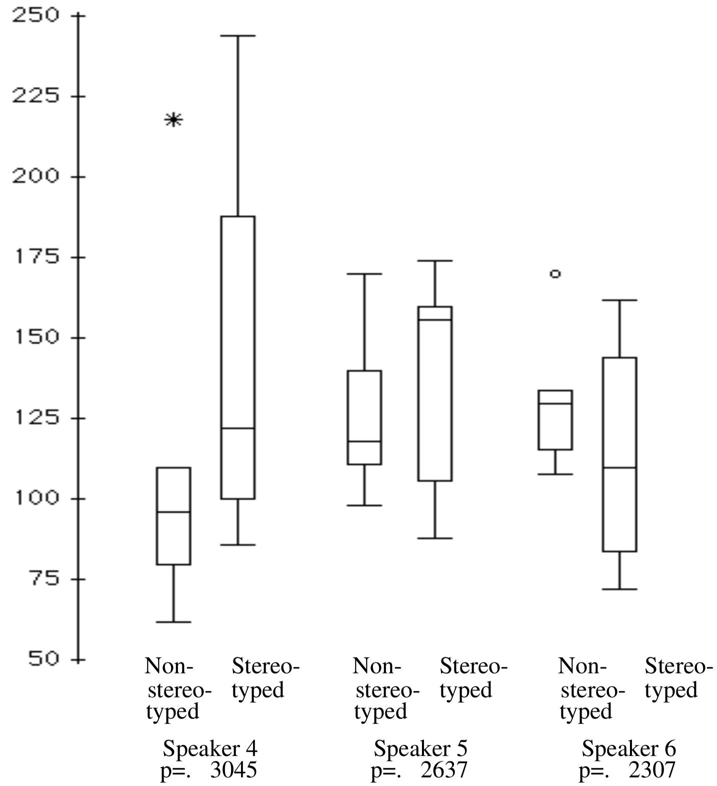


Figure 9: Durations for /#st/

For Speakers 4 and 5, the mean duration of /s/ in /#st/ is longer in stereotyped than in non-stereotyped speech, but this effect is massively insignificant. This finding is consistent with that of the first experiment; /#st/ seems not to participate in the lengthening found in /#sp/ and /#sk/. Speaker 4 appears to have very long /s/ durations in /#st/ overall, but these durations do not differ significantly between stereotyped and non-stereotyped speech.

3. Discussion

By finding quantitative support for the long-standing intuition that listeners can identify the sexual orientation of speakers strictly on the basis of phonetic cues, Gaudio set a research agenda for discovering what those cues may be. Since there are a great many dimensions along which speech can vary, this question is too large for any single study to answer. Both Gaudio and the present study have surveyed patches of the broad phonetic landscape.

At present, the tentative findings are as follows. The /s/-frication of /sp/ and /sk/ (at least as the onset of a word-initial stressed syllable) is lengthened in gay male stereotyped speech. /st/ appears not to be subject to this lengthening. Why /st/ should fail to participate in this pattern is not clear. Many possible explanations have been suggested by others after earlier presentation of the present work (e.g. perhaps the answer has to do with the fact that /s/ and /t/ have the same place of articulation, or that /t/ is coronal, etc.), but I presently have no basis for choosing among these possibilities.

Word-initial /l/ shows the same sort of systematic lengthening in stereotyped speech as is found for /sp/ and /sk/, but there are specific words for which this lengthening of /l/ is not found.

All of these findings are true for five out of the six subjects. Speaker 6, for whom none of these generalizations hold, is in his mid-fifties and is at least ten years older than the next youngest speaker. Subjectively, my impression is that Speaker 6 is "queenier" in his personality than any of the other speakers. Perhaps these observations might help toward an eventual explanation of Speaker 6's anomalous data, but it is not presently known what independent variable might predict a Speaker 6-type pattern.

Given the data currently available, the correct generalization would seem to be that a speaker may either employ or not employ /s/-lengthening in the production of a gay male stereotype; however, lengthening in /sp/ and /sk/ appear in tandem. Based on the limited number of speakers examined so far, it does not seem that a speaker can employ lengthening for /#sp/ but not /#sk/, nor /#sk/ but not /#sp/.

As is usually the case, a single answer raises a great many more questions. Is the artificial (electronic) manipulation of duration of /s/ and /l/ in the appropriate environments sufficient to induce a different listener judgment about the sexual orientation of the recorded speaker? What other segments seem to vary in this manner, and in what environments? Are other segments (e.g. syllable nuclei) compensatorily shortened in syllables where /l/ and /s/ are lengthened as described? What demographic variables correlate with a speaker's use or non-use of the described lengthening in the production of stereotyped speech?

Zwicky (in press) notes that it may be very difficult to determine the phonetic cues salient for sexual orientation identification of male speakers, since it is not necessarily the case that all speakers employ the same cues. By way of analogy, Zwicky cites the phonetic strategies used to disambiguate structures of the form *big cats and dogs,* where each speaker *emits* only one out of a set of several available cues to the intended parsing (e.g. pausing), but where all speakers *recognize* all of these strategies, even those which they do not personally employ. It could likewise be the case that the cues for sexual orientation identification are drawn by each speaker from such a set which all listeners recognize; in this case, it could prove very difficult to identify the set of salient cues. In the natural sciences, one should assume simplicity until the facts force complexity. However, the present findings seem to be consistent with the hypothesis that not all speakers employ the same cues in producing a gay male stereotype.

In future research, it would be of interest to investigate whether the lengthening of certain onsets is compensated by a shortening of other segments. For example, it could conceivably be the case that vowel nuclei are compensatorily shortened. This is an important question, because if compensatory shortening occurs, then an additional possible explanation for the informally observed "swoopiness" of gay men's F0 might be available. In particular, it is established that a difference in F0 contour (flat vs. rise-fall) can induce a difference in the perception of vowel length (van Dommelen, 1993)[8]. If it turns out that the converse is true (which would not be especially surprising), then one could reasonably expect that a compensatory shortening of a vowel would result in a *perceived* change in the F0 contour (a perceived higher peak when the peak is not in fact measured to be higher). This situation would be not be expected to be detectable by the analyses used in Gaudio's study.

Further, it has long been known that there are differences between adult males and females with regard to spectra of /s/; indeed, listeners are able to identify the sex of speakers from unvoiced fricatives played in isolation (Ingemann, 1968; Schwartz, 1968). While segmenting speech for the present study, I informally and subjectively observed that at least one of the subjects was pronouncing the /s/ segments in his stereotyped speech with a markedly exaggerated lisp. This observation is consistent with the widely made but (to my knowledge) quantitatively unverified claim that lisping is a cue which allows gay men to be identified. Empirical work will be needed to confirm

---

[8] The relationship here is actually rather complicated. It has been known for a long time that in the case of *isolated* vowels, an up-down F0 contour results in the vowel being perceived as longer than an isochronous vowel with a flat F0 contour. Van Dommelen (1993), however, found that this relation is exactly reversed when the vowel occurs within the context of a whole syllable. If the converse of this relationship turns out to be true, then the prediction would seem to be as follows: if vowels within syllables are shortened without a change in the initial, peak, and final values of F0, then the F0 contour of the shortened vowel will be perceived to be exaggerated.

or disconfirm this stereotype.

References

Gaudio, R. (1994) "Sounding Gay: Pitch Properties in the Speech of Gay and Straight Men." *American Speech* 69.1:30-57.

Ingemann, F. (1968) "Identification of the Speaker's Sex from Voiceless Fricatives." *Journal of the Acoustical Society of America* 44.4:1142-1144.

Schwartz, M. (1968) "Identification of Speaker Sex from Isolated, Voiceless Fricatives." *Journal of the Acoustical Society of America* 43.5:1178-1179.

Van Dommelen, W. (1993) "Does dynamic F0 increase perceived duration? New light on an old issue." *Journal of Phonetics* 21:367-386

Zwicky, A. In press. "Two Lavender Issues for Linguists," in Kira Hall and Anna Livia, eds., *Queerly Phrased.* Oxford Univ. Press.

Department of Linguistics
University of Pennsylvania
Philadelphia, PA 19104-6305
kurisuto@unagi.cis.upenn.edu