THE ROLE OF SOCIOINDEXICAL EXPECTATION IN THE PERCEPTION OF GAY MALE SPEECH

Dominique A. Bouavichith

University of Michigan
dombouav@umich.edu

ABSTRACT

During speech perception, listeners adjust expectations when given social information about a speaker; this effect has been demonstrated using many socially indexed phonetic cues. Previous sociophonetic studies have identified lengthened /s/ as one acoustic correlate of gay(-sounding) speech.

This eye-tracking study examines a) how information about a speaker's social identity affects listeners' time course of lexical activation and b) how listener experience mediates this process. Participants completed a categorization task of /s/-lengthened /CVs/-/CVsC/ minimal pairs in two blocks: the first presented no information about speaker sexuality; the second introduced auditory social primes indexing the speaker as gay.

Effects of social expectation were found as a function of listeners' experience with gay speech. High- and low-experience groups behaved similarly before social primes were introduced. As predicted, target fixations were delayed as a result of these primes, but only by high-experience listeners, demonstrating the expectation-mediating role of listeners' sociolinguistic experience.

Keywords: speech perception, sociophonetics, priming, social information, eye-tracking

1. INTRODUCTION

1.1. Socioindexical Expectation

During speech perception, listeners are presented with a rich phonetic signal, which provides linguistic information as well as information about the speaker's identity. The results of sociophonetic investigations have demonstrated that speakers can signal social group identification using fine-grained phonetic details, often below the level of speakers' consciousness [see 6].

Social information has also been shown to influence speech perception. Previous research has shown that social expectations about a speaker's gender [25], regional/national identity [18, 15], age [9], ethnicity [24], and sexual orientation [11, 16], among other factors [see 7 for a review] influence listeners' linguistic decisions. Many of these studies are grounded within exemplar frameworks, in which listeners use socially-coded exemplars to make predictions about upcoming speech.

In an experiment by Strand & Johnson [25], for example, listeners heard gender-ambiguous tokens on a continuum from 'shod' to 'sod' while looking at either a male or female face. Overall, seeing a female face elicited more 'shod' responses, and seeing a male face elicited more 'sod' responses. This reflects the difference in /ʃ-s/ category boundaries between men and women in production and shows that listeners are able to use social information (in this case, presented visually) to aid in speech perception.

Some of these studies consider the perception of speech over the time course of one word or one segment [5, 27, 15], which allows for fine-grained measures of processing that can test listeners' response to sociolinguistic information as it is given in real time.

1.2 Gay(-Sounding) Speech

Several acoustic correlates have been associated with gay speech; these include vowel space size [19], word-final stop release presence [20], phonation type [21], /s/ spectral quality [16, 29, 4, 14], pitch variation [8, 12, 28], and /s/ duration [23, 13, 12, 3].

This study uses /s/ duration as the primary sociophonetic variable of interest. In production, Linville found that, on average, gay men's /s/ durations are 17ms longer than straight men's, across many phonological contexts [13]. Additionally, Bouavichith found a similar pattern in production: when controlled for speech rate, queer men had significantly longer /s/ durations in /sVC/, /sCVC/, and /CVsC/ contexts [3]. This difference was not found in word-final singleton /s/ contexts (/CVs/).

In perception, Levon found that the combination of two socially encoded acoustic cues—wider pitch variation and lengthened /s/ segments—could reliably increase listeners' valuations that the speaker was gay [12]. When only one cue was used, however, valuations were not significantly altered. It seems, then, that multiple cues may be needed to make this social valuation.

The present investigation asks if a social cue—like those used in the socioindexical expectation literature—can be combined with an acoustic cue to

inform listeners' perception patterns, as measured using eye-tracking.

2. METHODS

2.1. Stimuli

Target auditory stimuli consisted of ten monosyllabic minimal triplets: (i) /s/-final, (ii) /s-stop/ cluster-final, and (iii) stop-final (e.g., 'bass', 'bask', 'back'). Stimuli were produced by a straight, male, native speaker of American English.

A single token of each target word was chosen for the test stimuli. Based on gay speakers' /s/ duration rates from Bouavichith [3], waveform editing techniques were applied in Praat [2] to create three lengthened, natural-sounding /s/ durations—short, mid, and long—for each token type (cluster, singleton). These durations were constant across all stimuli (e.g., all 'short' condition cluster stimuli (bask, mosque, mast, etc.) had equal /s/ durations).

These /s/ tokens all fell within the range of /s/ durations described as characteristic of gay speech in prior investigations. The /s/ durations, thus, were longer in singleton contexts than in cluster contexts. Three levels were used to ensure variation in the /s/ durations heard by listeners, in order that duration would not become predictable over the course of the experiment, while approximating a somewhat representative sample of speech.

Stimuli were cross-spliced to ensure that participants heard identical acoustic information for each target word up to the onset of /s/ (e.g., bass, bask) or a word-final singleton stop (e.g., back). To ensure that /s/ was identical in the /CVs/ and /CVsC/ conditions, but that the /s/ in /CVsC/ was not coarticulatorily affected by the word-final stop, the /s/ for the /CVsC/ stimuli was spliced from the /CVs/ condition; this ensured that there were no spectral, and only temporal, differences between cluster and single /s/ conditions. Given the existing literature on spectral effects on the socioperception of gay speech, extra care was taken to avoid this confound.

The visual stimuli consisted of greyscale line drawings, corresponding to each of the 30 target words.

2.2. Procedure

2.2.1. Participants

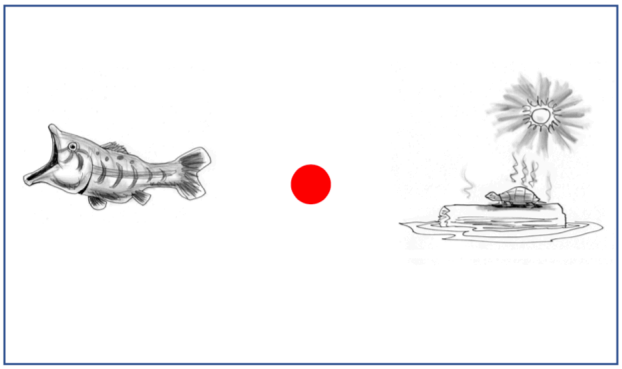
Thirty-four male undergraduate students, all native speakers of American English, participated in this study. Two participants were excluded (poor overall performance and technical difficulties); data from 32 participants were included in the analyses.

2.2.2. Stimuli and procedure

Participants were presented with auditory stimuli over headphones, and their eye movements were recorded using a remote monocular eye-tracker (EyeLink 1000 Plus, SR Research).

The experiment consisted of five blocks of 48 stimuli. During trials, participants were presented with two images (target and competitor), corresponding to two of the three words in each minimal triplet. There were three trial types: singleton /CVs/ vs. stop (e.g., 'bass'/'back'), cluster /CVsC/ vs. stop (e.g., 'bask'/'back'), and singleton /CVs/ vs. cluster /CVsC/ (e.g., 'bass'/'bask'). Image positions were counterbalanced across trials. The critical trials used for analysis were the singleton (/CVs/) vs. cluster (/CVsC/) contexts. A sample screen of a trial, in which 'bass'/'bask' was the target/competitor pair, can be seen in Figure 1. There were 120 critical trials throughout the experiment.

Figure 1: Sample screen of visual stimuli ('bass'/'bask').



Each trial lasted approximately 10 seconds and consisted of the following sequence, in which all auditory stimuli were produced by the same speaker:

1. Both images were shown with the accompanying audio: "Look at each image."
2. After 2000ms, a filled red circle appeared in the center of the screen with the audio "Look at the middle dot."
3. While the dot was still on the screen, participants then heard "My [relation] told me: 'Now, look at ____."
4. Immediately before the auditory target word was heard, the red dot disappeared from the screen. Recording continued for 3000ms following the disappearance of the red dot.

The priming of social information took place in the second half of the experiment, with a manipulation of the [relation] term. In the first half of the experiment, the [relation] in the presentation phrase was socially neutral: participants heard 'brother', 'mother',

'father', and 'friend'. Halfway through the experiment, two additional relationship terms—'boyfriend' and 'partner'—were added to prime the male speaker's sexual orientation. Each target word in the critical trials was presented six times: three (three length conditions) in the pre-exposure phase and three in the post-exposure phase. This resulted in 30 cluster stimuli and 30 singleton stimuli in each half of the experiment.

To gauge participants' experience with gay people and gay speech, participants also completed a demographic survey that asked them to provide their sexual orientation, as well as a rating of their connectedness to the LGBTQ community, on a ten-point scale (to serve as a proxy for language experience). These data were later recoded in binary terms (queer/non-queer and high-/low-experience). This survey was administered at the end of the testing session.

2.3. Measures and Predictions

The eye movements of each participant were monitored throughout the trial, and a critical interval—from 200ms to 1000ms after the onset of the /s/ in each target—was used for analysis. The dependent variable was the proportion of correct fixations over time. Trials were analyzed using exposure phase as a grouping factor (before and after social primes were introduced).

After social exposure, participants who are more experienced with gay speech are expected to use this experience to anticipate longer /s/ durations and delay their looks to a target. Low-experience participants, however, are not expected to show change between exposure phases.

Because phrase-final singleton /s/, like all phrase-final segments, is generally lengthened and therefore possibly less socially marked in /CVs/ contexts, greater pre- vs. post-exposure differences are predicted for cluster conditions (i.e., when listeners heard 'bask' in a critical trial) than for single /s/-final conditions (i.e., 'bass'). This is supported by production data, which shows a significant difference between gay and straight men's /s/ durations in /CVsC/ but not /CVs/ words [3]. Across the three /s/ duration conditions, the highest degree of ambiguity between phrase-final lengthening and socially motivated lengthening is predicted for the most lengthened /s/ conditions. Therefore, it is predicted that the long /s/ condition will elicit the strongest effect of socioindexical expectation.

3. RESULTS

3.1. Demographic Results

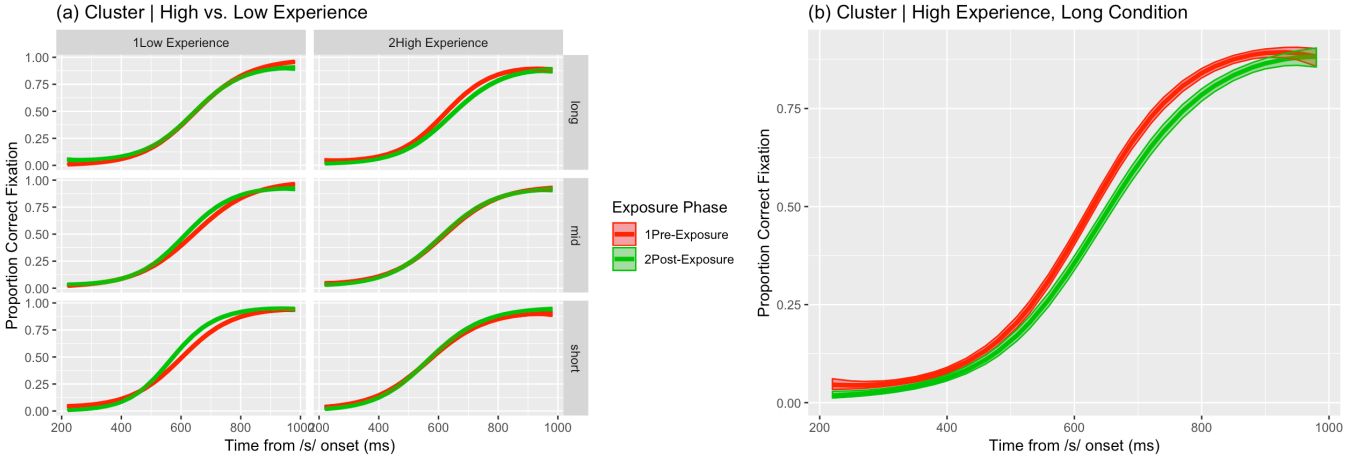
Of the 32 participants included in analysis, five reported a sexual orientation other than 'straight'. Due to high imbalance between groups, this was not used as a factor in analysis.

The experience metric ranged from 1 to 8, with a mean of 4.26. This scale was converted to a binary high- vs. low-experience metric, where high ≥ 5.

3.2. Eye-Tracking Results: Target Fixations

Measures of accuracy (for the singleton vs. cluster trials) during the critical interval period were smoothed using 20ms temporal bins before being modeled. Generalized Linear Models were used to fit the fixation data using the *lme4* package [1] in R [21], with the proportion of correct target fixations as the dependent variable, and the independent variables of b-spline-modeled time, /s/ duration condition, experience (binary), exposure phase, and their interactions. Two models with this structure were run:

Figure 2: (a) GLM-predicted results for all cluster accuracy data, showing exposure phase, experience level, and stimuli duration. Rightward shifts of the post-exposure curves (green) indicate delay to target fixation, indicating listeners' use of social information to wait for word-final silence or stop release. **(b)** Enlarged plot of high-experience listeners in the long (most ambiguous) /s/ condition, showing most extreme rightward shift.



1) singleton (/CVs/) auditory targets and 2) cluster (/CVsC/) auditory targets.

3.2.1. /CVs/ Trials

Significant main effects were found for experience (z = 2.98, p = 0.003), and, as expected, for temporal measures (b-spline-modeled time, /s/ length condition). Low-experience listeners were generally slower to fixate on target images, but there was not a significant effect of exposure phase (z = 1.41, p = 0.123), nor was there a significant interaction between exposure phase and experience (z = 0.51, p = 0.612). This is consistent with previous production research showing smaller differences in duration for word-final singleton /s/ between sexual orientation groups [3].

3.2.2. /CVsC/ Trials

The GLM conducted on the proportion of correct target fixations in the cluster trials showed significant effects of exposure phase (z = -5.12, $p < 0.001$), of temporal measures (b-spline-modeled time, /s/ length condition), and of the interaction between exposure phase and experience level (z = 2.10, p = 0.036). Unlike the singleton model, there was no main effect of listener experience (z = -0.75, p = 0.454). Visualizations of the model predictions are given in Figure 2. Figure 2a shows that low-experience listeners fixate on the target image *more quickly* in the post-exposure phase, whereas high-experience listeners show no change or a delay in fixations, depending on the /s/ duration condition. A significant three-way interaction of experience, exposure, and /s/ duration was found (long vs. mid /s/ duration: z = -6.17, $p < 0.001$).

4. DISCUSSION

4.1. Summary of Results

As predicted, the effect of the social exposure condition only emerged in the /CVsC/ condition. Predicted fixation patterns—delays for high-experience participants in the long /s/ condition—were only found for cluster trials; this delay is shown in Figure 2b, where the post-exposure curve is shifted toward the right (i.e., correct target fixations are reliably delayed). The leftward shift of the post-exposure curves among low-experience listeners is tentatively explained by a task-specific learning effect: low-experience listeners are getting more familiar with /s/ durations in the stimuli and are, therefore, responding more quickly; this learning effect appears to be counteracted by socioindexical expectation among the high-experience listeners. It is also possible that both groups *are* using socioindexical information, but in different ways. Further exploration is needed to understand this relationship. Additionally, the longest stimulus conditions were found to have the longest fixation latencies, irrespective of experience group.

4.2. General Discussion

The lack of an effect of exposure phase in /CVs/ trials is consistent with production patterns [3], where no significant difference was found between gay and straight speakers in singleton-/s/-final tokens. This lack of difference in production suggests a lower likelihood in the social encoding of gay speech in this phonological environment, specifically based on /s/ duration. Because tokens were phrase-final, variation in phrase-final lengthening likely affected the social categorizability of these tokens, and no interaction of experience and exposure phase was found.

The observed effect of experience in the /CVs/ model, however, may point to more general differences between experience groups. It is possible that listeners come to the task with different default assumptions about the speech patterns associated with a given identity based on their own sociolinguistic experience.

The much stronger effects in the /CVsC/ model—specifically the interaction of listener experience and social exposure phase—show that the perceptual time course was in fact affected in the expected way. This result provides strong evidence of socioindexical expectation as a function of listener experience.

Resonance exemplar models of social-linguistic integration have been proposed, in which utterances are stored and encoded with social information [10, 26]. Within these models, activation of exemplars of socially encoded utterances is strengthened when perceived utterances match listeners' expectations of how social identity is represented in the speech signal. These expectations reflect higher density of stored episodic traces within a given listener's socio-acoustic cloud. Therefore, a listener who has more experience with any given sociolinguistic variety—in this case, gay speech—will have a stronger activation of these utterances. With increased exemplars, listeners can form these socio-perceptual expectations to inform—and in this case delay—linguistic decisions.

5. ACKNOWLEDGMENTS

I thank audiences at the University of Michigan, the Acoustical Society of America, and New Ways of Analyzing Variation for helpful comments and thoughtful critiques.

6. REFERENCES

[1] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1), 1-48.

[2] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5, 341-345.

[3] Bouavichith, D.A. 2017. Cross-contextual consistency of /s/ length and spectral quality in gay men's speech (poster). *174th Meeting of the Acoustical Society of America*, New Orleans, LA.

[4] Campbell-Kibler, K. 2011. Intersecting variables and perceived sexual orientation in men. *American Speech* 86(1), 52-68.

[5] Dahan, D., Drucker, S.J., Scarborough, R.A. 2008. Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition* 108, 710-718.

[6] Foulkes, P., Docherty, G. 2006. The social life of phonetics and phonology. *Journal of Phonetics* 34(4), 409-438.

[7] Drager, K. 2010. Sociophonetic variation in speech perception. *Language and Linguistics Compass* 4(7), 473-480.

[8] Gaudio, R. 1994. Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech* 69, 30–57.

[9] Hay, J., Warren, P., Drager, K. 2006b. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34, 458-484.

[10] Johnson, K. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34, 485-499.

[11] Levon, E. 2006. Hearing gay: Prosody, interpretation and the affective judgment of men's speech. *American Speech* 81(1), 56-78.

[12] Levon, E. 2007. Sexuality in context: Variation and the sociolinguistic perception of identity. *Language in Society* 36(4), 533-554.

[13] Linville, S. 1998. Acoustic correlates of perceived versus actual sexual orientation in men's speech. *Pholia Phoniatrica et Logopaedica* 50, 35– 48.

[14] Mack, S., Munson, B. 2012. The association between /s/ quality and perceived sexual orientation of men's voices: implicit and explicit measures. *Journal of Phonetics* 40, 198–212.

[15] McGowan, K. 2015. Social expectation improves speech perception in noise. *Language and Speech* 58(4), 502-521.

[16] Munson, B. 2007. The acoustic correlates of perceived sexual orientation, perceived masculinity, and perceived femininity. *Language and Speech* 50, 125-142.

[17] Munson, B., McDonald, E.C., DeBoe, N.L., White, A.R. 2006. The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics* 34, 202-240.

[18] Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18(l), 62-85.

[19] Pierrehumbert, J.B., Bent, T., Munson, B., Bradlow, A.R., Bailey, J.M. 2004. The influence of sexual orientation on vowel production (L). *Journal of the Acoustical Society of America* 116(4), 1905-1908.

[20] Podesva, R. 2004. On constructing social meaning with stop release bursts. *Sociolinguistics Symposium* 15, 1-5.

[21] Podesva, R.J. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics* 11(4), 478-504.

[22] R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

[23] Rogers, H., Smyth, R., Jacobs, G. 2000. Vowel and sibilant duration in gay-and straight-sounding male speech. *First International Gender and Language Association Conference*, Stanford, California.

[24] Staum Casasanto, L. 2008. Does social influence sentence processing? *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, Washington, D.C.

[25] Strand, E.A., Johnson, K. 1996. Gradient and visual speaker normalization in the perception of fricatives. In: Gibbon, D. (ed), *Natural Language Processing and Speech Technology: Results of the 3rd KOVENS Conference.* Berlin: Mouton de Gruyter, 14-26.

[26] Sumner, M., Seung, K., King, E., McGowan, K. 2014. The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psychology* 4, 1015.

[27] Trude, A.M., Brown-Schmidt, S. 2012. Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes* 27(7-8), 979-1001.

[28] Wilbanks, E. 2013. Pitch and VOT as Factors in the Perception of Sexual Identity and Masculinity in Male Speech. *New Ways of Analyzing Variation 42,* Pittsburgh, PA.

[29] Zimman, L. 2013. Hegemonic masculinity and the variability of gay-sounding speech. *Journal of Language and Sexuality* 2(1), 1-39.