Perception of Spanish in contact:
Bolivian listeners' expectation and awareness of socioindexical variation.

There is mounting evidence that socioindexical expectations can influence listener performance on a variety of behavioral measures (Drager, 2010; Foulkes, 2010). Manipulating listeners' beliefs about the age, gender, sexual orientation, race, etc. of the talker can lead to behavioral responses which suggest that listeners process the speech signal differently given different social expectations. Perception of particular phonetic cues can be altered in response to a primed social group. In other words, listeners appear to dynamically alter the attentional weights associated with particular phonetic cues in response to manipulated social and indexical expectations (Pierrehumbert, 2002). This alteration has been shown to occur both with cues that reflect actual usage (e.g. Schulman, 1983; Hay et al., 2006) and stereotypical usage (e.g. Mack and Munson, 2012). Socioindexical perception effects suggest that speech perception proceeds, not by winnowing away noise to arrive at a core, intended signal, but by exploiting real patterns of informative variation to impose structure upon the phonetic signal.

This position, that systematic patterns of variation provide essential support for listeners, is certainly not without precedent. Sumner (2011) finds that "variation is both necessary and beneficial" for English listeners overcoming a gross categorical mismatch in French-accented patterns of voice onset time (what Best (1995) classifies as the single category assimilation of two variants). Whalen (1984) demonstrates that listeners are sensitive to subcategorical acoustic mismatches below the level of experimenter awareness. The gating task of Lahiri and Marlsen-Wilson (1991) finds that listeners hearing portions of CV and CṼ stimuli report hearing a CVN, rather than CVC, target when vowel nasalization is present. Beddor et al. (ming) report eye-tracking evidence that coarticulatory cues speed the time course of lexical activation –with listeners in a forced-choice visual world task taking immediate advantage of available cues to nasalization. Beddor (2009), citing decades of research on the perceptual consequences of coarticulation, argues that coarticulation provides listeners with informative variation which gives structure to the acoustic signal and simplifies the task of perception. Individual listeners assign different weights to particular coarticulatory cues consistent with their experience of the variability and usefulness of those cues in understanding speech. This reweighting of cues leads to the actuation of sound change without the innocent misperception stage of Ohala (1981) – listeners with reweighted cues simply attribute identical signals to different underlying sources than listeners for whom this reweighting has not occurred.

For Beddor, this reweighting is particular to coarticulation. However, if listeners use patterns of socially-informative variation available in the speech stimulus and if this invokes a set of specific phonetic expectations that covary with contextual cues then manipulated social expectations –or socially-motivated cue weightings– should similarly alter listener perceptions and set the stage for the initiation of sound change. Sumner and Samuel (2009) have shown that listener experience with a particular accent –Long Island English– predicts the usefulness of semantic priming by voices with that accent. Similarly, Szakay et al. (2012) used a cross-language and cross-dialect priming task to demonstrate that variable sociophonetic cues can facilitate translation priming. It seems clear that listeners have –and can use– knowledge of particular social variants during perception. The present study uses both experimental and ethnographic methods to identify the links between social categorization and the phonetic cues to this categorization. We test the ability of listeners to access information below the level of conscious awareness as they use social information to guide perception of synthesized stimuli as members of linguistic categories. This paper sheds light both on listeners' sensitivity to detailed patterns of socially-informative phonetic variation and on listeners awareness of these patterns as potential predictors of linguistic variation.

We examine the relationship between speech perception and identity categories in an area of language contact between Spanish and Quechua in central Bolivia. We use an AXB psychoacoustic discrimination task to test the hypothesis that manipulating expectation of speaker identity as a Quechua-oriented speaker of Spanish or as a Spanish-oriented speaker of Spanish will shift listener category boundaries. When Quechua-dominant speakers target /e/ and /o/ sounds in Spanish, they often produce /ɪ/ and /ʊ/, which Spanish-dominant speakers then interpret as /i/ and /u/ respectively. This phenomenon is known as 'motosidad' and is highly metapragmatically salient, giving rise to hypercorrections, which are also parodied. A second contrast, stemming from the same linguistic factors, concerns /a/-/æ/ variation. While this variation also occurs among Quechua-dominant speakers, it is not metapragmatically salient.

20 Spanish-speaking listeners participated in a within-subjects matched-guise AXB discrimination task with tokens from synthesized /i/-/e/ (pisa-pesa or pica-peca) and /o/-/u/ (moda-muda or soda-suda) continua presented over headphones at the Bolivian field site. Half of the participants initially believed they were hearing a Spanish-dominant speaker from Santa Cruz while the other half initially believed they were listening to a Quechua-dominant speaker from Cochabamba. The experimenter asked debriefing questions after this first presentation and then reversed the speaker guise and presented a second block of synthesized tokens from the same

voice.

(a) Initial Guise Presentation

(b) Second Guise Presentation
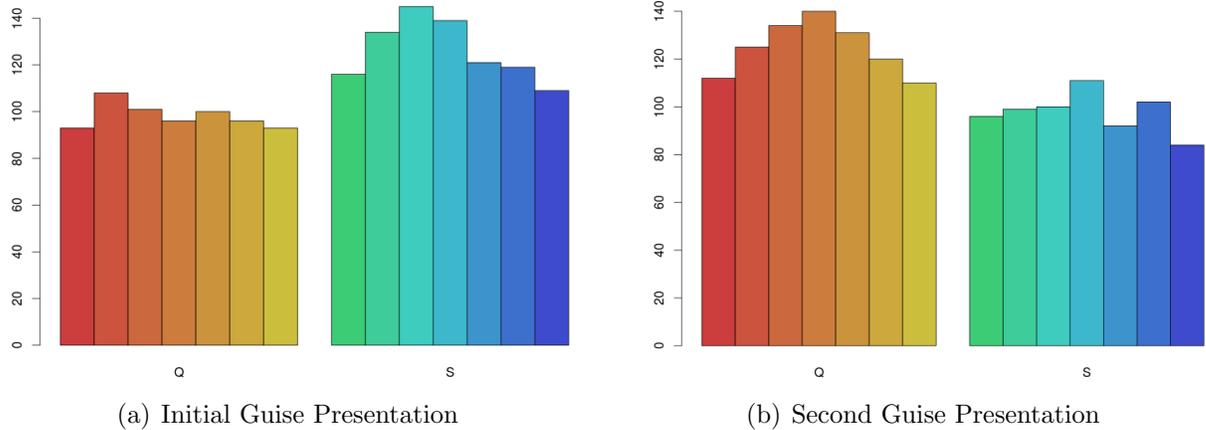
Manipulating expectation of speaker identity shifted listener category boundaries for /e/-/i/ and /o/-/u/, but only for the *initial* guise presentation (β = -0.229; $p < .05$; see Table 1) as shown in figure 1a. Listeners initially presented with the Spanish-dominant guise (blue bars) show a clear category boundary between the extremes of the continua with lowest disciminability toward the center of the continua. Listeners initially presented with the Quechua-dominant guise (red bars), on the other hand, show a much flatter discrimination function suggesting less distinct perceptual categories.

	Coef β	SE(β)	z	p
(Intercept)	0.712	0.141	5.063	**<0.0001**
guise	−0.012	0.032	−0.405	> 0.6
presentationOrder	−0.009	0.032	−2.777	> −0.7
I(guise|presentationOrder)	−0.229	0.099	−2.319	**<0.05**

Table 1: **response accuracy in terms of guise x presentation order**

Perhaps unsurprisingly, the attempt to present the same speaker's voice under the alternative social guise appears to have had no effect on the listeners' discrimination patterns between first and second presentations (figure 1b). Category boundaries for the vowels in each continuum emerge in the first guise presentation and are then retained across the switch from Spanish to Quechua or from Quechua to Spanish guise conditions. Crucially, however, and in contrast to these empirical results, interviews

conducted immediately following the discrimination task give every indication that participants believed the guise switch had occurred –commenting on everything from the new speaker's pronunciation to her relative education level.

Data were analyzed using the lme4() function (Bates et al. 2011) of the R statistical analysis environment (R Development Core Team, 2011). Using a generalized linear mixed model with binomial errors and a logit link function in which response accuracy (correctly identifying which item, A or B, matched the reference item X) was the dependent variable and the interaction term of guise level (Spanish-dominant or Quechua-dominant) by guise presentation order (whether the Spanish-dominant or Quechua-dominant guise was presented first) was modeled as a fixed effect. Participant and item were included as random effects with varying intercepts (and fixed slopes). Results of this linear model are shown in Table 1.

This evidence suggests that listeners have detailed phonetic expectations about how members of certain social categories speak. It is likely that listeners rely on these expectations most heavily when encountering a new or unfamiliar speaker who is clearly marked for social category in some other way. A second conclusion, however, is that once expectations are associated with a particular voice these this association is not easily changed –at least at the unconscious level. Finally, in explicit commentary, speakers rely more on social stereotypes than on the acoustic signal and, at this conscious level, appear more willing to disregard the acoustic evidence in favor of other sources of socioindexical information.

The present study has a number of possible implications for the actuation of sound change. Listeners appear to be accumulating evidence of the relationship between social category and acoustic cue without necessarily being aware of the variation or of the linkage. These listeners then appear to use their detailed phonetic expectations when perceiving a new speaker/voice. This result suggests a social analog to Beddor (2009)'s coarticulatory path to sound change. Listener expectations include detailed phonetic knowledge about the speech of the social categories associated with that speaker. If different listeners attend to these social cues differentially, perhaps by identifying with the social category (Labov, 1994), it could initiate a cue reweighting ultimately leading to sound change.

The results of the changed guise demonstrate that, despite the detailed phonetic knowledge evident in the first presentation, listeners have neither a high degree of awareness about these detailed phonetic patterns nor an ability to discuss the relationships between acoustic cues and social categories. This second result suggests a low degree of agency on the part of the listener in the actuation of sound change. However, this is not the innocent misperception of Ohala (1981) or, indeed, misperception at all. Instead, the results of the present study add to converging evidence

(Niedzielski, 1999; Hay et al., 2006; Drager, 2009; McGowan, 2011; Sumner, 2011, etc.) that speech perception is exquisitely sensitive to listeners' expectation of social variation. We will not go so far as to say that the acoustic signal is unimportant, but it is certainly meaningful only as part of a pattern of features, both linguistic and non-linguistic, that conjure a set of expectations on the part of listeners.

References

Beddor, P. (2009). A coarticulatory path to sound change. *Language*, 85(4):785–821.

Beddor, P., McGowan, K., Boland, J., Coetzee, A., and Brasher, A. (forthcoming). The perceptual time course of coarticulation.

Best, C. (1995). A direct realist perspective on cross-language speech perception. In Strange, W. and Jenkins, J., editors, *Cross-language speech perception*, pages 171–204. York Press, Timonium, MD.

Drager, K. (2009). *A Sociophonetic Ethnography of Selwyn Girls' High*. PhD thesis, University of Canterbury, Canterbury, NZ.

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

Foulkes, P. (2010). Exploring social-indexical variation: a long past but a short history. laboratory phonology. *Journal of Laboratory Phonology*, 1.

Hay, J., Warren, P., and Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34:458–484.

Labov, W. (1994). *Principles of Linguistic Change: Volume I: Internal Factors*. Blackwell Publishers, Cambridge, MA.

Lahiri, A. and Marlsen-Wilson, W. (1991). The mental representation of lexical form: a phonological approach to the recognition lexicon. *Cognition*, 3:245–294.

Mack, S. and Munson, B. (2012). The influence of /s/ quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp' stereotype. *Journal of Phonetics*, 40(1):198–212.

McGowan, K. B. (2011). *The Role of Socioindexical Expectation in Speech Perception*. PhD thesis, University of Michigan, Ann Arbor, MI.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, 18(1):62–85.

Ohala, J. J. (1981). The listener as a source of sound change. In Masek, C. S., Hendrick, R. A., and Miller, M. F., editors, *Papers from the Parasession on Language and Behavior*, pages 178–203. Chicago Linguistics Society, Chicago.

Pierrehumbert, J. B. (2002). Probabilistic phonology: Discrimination and robustness. In Bod, R., Hay, J., and Jannedy, S., editors, *Probability Theory in Linguistics*. MIT Press.

R Development Core Team (2011). R: A language and environment for statistical computing. ISBN 3-900051-07-0.

Schulman, R. (1983). Vowel categorization by the bilingual listener. *PERILUS Working Papers*, III:81–99.

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, 119:131–36.

Sumner, M. and Samuel, A. G. (2009). The role of experience in the processing of cross-dialectal variation. *Journal of Memory and Language*, 60:487501.

Szakay, A., Babel, M., and King, J. (2012). Sociophonetic markers facilitate translation priming: Maori english goat a different kind of animal. *University of Pennsylvania Working Papers in Linguistics*, 18(2).

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics*, 35:49–64.