Talking, like, a Valley Girl? Online Processing of Sociolinguistic Cues

Daisy Leigh, Judith Degen, Robert J. Podesva (Stanford University)

Listeners form impressions about a speaker's social persona both from *what* they say and *how* they say it [7,8]. How and when do we do this? Sociolinguistic work has long shown that the social meanings associated with specific phonetic variants ('cues') contribute to listeners' perceptions of speaker persona in offline judgments: for example, speakers may sound more 'casual' and 'unprofessional' when they use *-in'* rather than *-ing* (e.g., *talkin'* vs. *talking*) [4,9]; or more 'excitable' or 'like a Valley Girl' when they use High Rising Terminals (HRT, aka 'uptalk') rather than declarative prosody [13]. But the effects of these cues are not fixed or absolute; *-in'* might indicate an 'unprofessional' persona in some voices but not others, for example [3,10]. This suggests that sociophonetic cues compete with other information in the speech signal; listeners must integrate the meaning contribution of sociophonetic cues with all the other social impressions that arise when hearing someone talk. As yet, little is known about how or when this happens: while most psycholinguistic work investigating the online processing of linguistic cues has focused on listeners' inferences about upcoming linguistic *material* (e.g., [1,6,11]), very little attention has been paid to listeners' unfolding inferences about the *speaker* (though see e.g., [2,7,14]). We conducted two eye-tracking studies to investigate listeners' online and offline uptake of two sociophonetic cues: *in'* (Exp.1) and HRT (Exp.2).

Design: In a 2AFC visual world paradigm, participants heard a stimulus and selected the speaker they thought produced it. The two speakers were representations of different personas: a Tough and a Valley Girl (Fig.1). We measured participants' persona selections and eye movements. On critical trials, participants in Exp.1 (N=160) heard *-in'* and *-ing* cues; those in Exp.2 (N=152), HRT and declaratives. Participants heard stimuli produced by four different voices in each experiment. Experiments were conducted online on Prolific, using Webgazer.js [10] to capture gaze data. **Predictions:** The Tough and Valley Girl images were normed to ensure they captured similar meanings to those reported for *-in'* (e.g. 'chill, unprofessional') vs. HRT (e.g., 'excited, feminine'). We therefore expected listeners to look towards, and select, Toughs more often after hearing *-in'* (vs. *-ing*) in Exp.1, and less after hearing HRT (vs. declaratives) in Exp.2. **Results:** Both cues modulated participants' offline judgements of speaker persona in the predicted directions, regardless of how Tough or Valley Girl the four different voices sounded overall (Fig.3 and 5). They also modulated online behavior: from the 800-900ms window after cue onset onwards, participants were more likely to look at the Tough image if they had heard *-in'* rather than *-ing*. In Exp.2, participants were significantly *less* likely to look at the Tough image after hearing HRT rather than a declarative, from the 1000-1100ms window onwards. *Overall* Tough/Valley bias for each voice was also reflected in looking patterns: e.g., in Exp.1, participants initially looked more to the Valley Girl persona when listening to Voice 4, but the presence of the *-in*' cue biased them towards the Tough interpretation (Fig. 4). **Discussion:** Our results suggest that participants processed both cues probabilistically by weighing the meaning contributions of each against their existing expectations about the speaker. Online cue uptake was observed much later than the 200ms typically allocated to executing signal-driven eye movements [1]. Given the sparsity of existing work, we can only speculate on the reasons: it is possible that listeners simply take longer to process phonetic cues' social meanings than their purely referential ones. Alternatively, these cues may be weak or less reliable cues to social identity; stronger/more reliable cues may result in faster online integration. Despite the delayed online effects, our results qualitatively (Figs.3-6) point to HRT having stronger biasing effects on interpretation than *-in'*, indicating that these cues vary with respect to their relative social informativity. These considerations raise exciting questions regarding the role of cue strength, reliability, and timing in the online integration of social and denotational information. We consider the current findings a promising starting point for future empirical work examining the online processing of sociophonetic cues.

Fig.1: Example Persona Images

Fig.2: Example stimuli

	Exp.1		Exp.2
-in'	I'm talkin' about the beam.	HRT	I'm talking about the **beam**↗
-ing	I'm talking about the **beam**.	Declarative	I'm talking about the **beam**.

All critical items took the form "I'm talking about the *x*", where *x* is a monosyllabic word. Participants heard 16 different items: 8 *-in'* items, 8 *-ing* in Exp.1, 8 HRT and 8 Declarative items in Exp.2. Stimuli were identical across conditions other than the differences outlined in Fig.2, and the same stimuli were used for the *-ing* and Declarative conditions in Exps. 1 and 2. We used existing utterances from the NSP corpus [5] and manipulated them to include the cues of interest.

Statistical details

For persona selections, we fit logistic regression models predicting log-odds of selecting a Tough persona given the presence of an *-in'* (Exp.1) or HRT cue (Exp.2). For both models, there was a main effect of cue (β=0.57, p<0.001 in Exp.1, β=-0.80, p<0.001 in Exp.2). For eye-tracking data, we fit logistic regression models predicting log-odds of looking at the Tough (vs. Valley Girl) in each 100ms window after cue onset, given the cue heard. (We took 200ms either side of cue onset as a baseline period against which looks in subsequent windows were compared). In Exp.1, the earliest window that the presence of -in (vs. ing) predicted a significant increase in Tough looks was 800-900ms (β=0.29, p<0.05). In Exp.2, the earliest window where HRT predicted a significant decrease in Tough looks was 1000-1100ms (β=-0.34, p<0.05). For all models, we included the maximal random effects structure justified by the data.

Figs.3-6 are ordered by 'Toughest' to most 'Valley Girl' sounding speaker; in both experiments, the voices were heard in random order.
Error bars represent bootstrapped 95% confidence intervals. In Fig.4 and 6, black verticals line represent cue onset, and pink lines, audio offset.

Fig.3: Exp.1, Persona selections

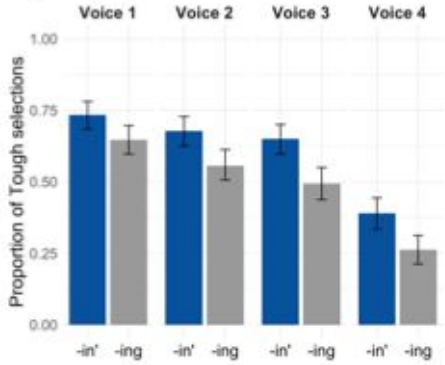


Fig.4: Exp.1, Eye-tracking data

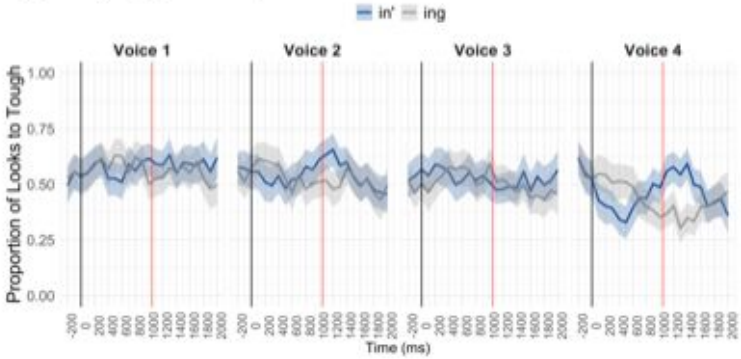


Fig.5: Exp.2, Persona selections

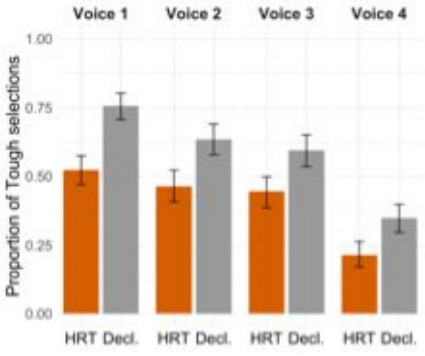
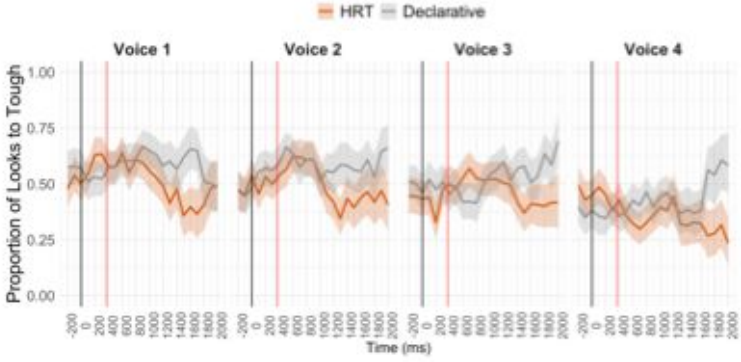


Fig.6: Exp.2, Eye-tracking data



References: **[1]** Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language.* **[2]** Austen, M. & Campbell-Kibler, K. (2020). Eye-tracking for sociolinguistic perception. Poster at *Annual Meeting of the Linguistic Society of America.* **[3]** Campbell-Kibler, K. (2008). I'll be the judge of that: Diversity in social perceptions of ING. *Language in Society*. **[4]** Campbell-Kibler, K. (2009). The nature of sociolinguistic perception. *Language Variation and Change.* **[5]** Clopper, C. G., & Pisoni, D. B. (2006). The Nationwide Speech Project: A new corpus of American English dialects. *Speech Communication.* **[6]** Degen, J. & Tanenhaus, M.K. (2016). Availability of Alternatives and the Processing of Scalar Implicatures: A Visual World Eye-Tracking Study. *Cognitive Science.* **[7]** D'Onofrio, A. (2018). Controlled and automatic perceptions of a sociolinguistic marker. *Language Variation and Change.* **[8]** D'Onofrio, A. (2018). Personae and phonetic detail in sociolinguistic signs. *Language in Society.* **[9]** Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M. & Nagy, N. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics* **[10]** Levon, E. & Fox, S. (2014). Social Salience and the Sociolinguistic Monitor: A Case Study of ING and TH- fronting in Britain. *Journal of English Linguistics.* **[11]** McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomics Bulletin Review.* **[12]** Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable Webcam Eye Tracking Using User Interactions. *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI).* **[13]** Tyler, J.C. (2015). Expanding and Mapping the Indexical Field: Rising Pitch, the Uptalk Stereotype, and Perceptual Variation. *Journal of English Linguistics.* **[14]** Weissler, R. E. & Brennan, J. R. (2020) How do Listeners Form Grammatical Expectations to African American Language?, *University of Pennsylvania Working Papers in Linguistics.*