CAMBRIDGE
UNIVERSITY PRESS

The role of internal constraints and stylistic congruence on a variant's social impact

Charlotte Vaughn

University of Oregon & University of Maryland, USA
Corresponding author. E-mail: cvaughn@uoregon.edu

Abstract

In natural conversation, multiple factors likely impact the social force of a sociolinguistic variant, yet researchers have tended to examine individual factors in isolation. This paper considers two underexamined factors together—the role of a variable's internal constraints and the role of stylistically congruent surrounding speech—to understand their combined influence on how a single variable's realization is socially interpreted. Focusing on English variable (ING), two accent rating experiments used stimuli varying the grammatical category of (ING) words and varying the stylistic congruence (natural sentences versus spliced stimuli) between (ING) realization and sentence frames. Results indicate that listeners showed sensitivity to (ING)'s internal constraints but only when the congruence between (ING)'s realization and other cues was not disrupted by using spliced stimuli. These findings suggest that internal constraints and stylistic congruence play a role in social signaling, and have methodological implications for the use of splicing.

Keywords: social evaluation; internal constraints; stylistic congruence; expectations; splicing

A linguistic form's social meaning is not uniform across situations but rather is co-constructed among the speaker, the listener, and the context. Listeners' experiences with linguistic forms, speakers, contexts, and surrounding language ideologies accumulate in their representations, and a given form's social impact results from the interplay between these factors (e.g., Eckert, 2012; Ochs, 1992; Podesva, 2008; Zhang, 2005). In the past several decades, experimental studies have examined this interplay directly, finding, for example, that how frequently a variant is used (e.g., Labov, Ash, Ravindranath, Weldon, Baranowski, & Nagy, 2011), and how likely the variant is in a particular social (e.g., Campbell-Kibler, 2009) and linguistic (e.g., Bender, 2005) context can affect how a speaker will be socially evaluated for their use of that variant. Moreover, clusters of variants of other variables occurring alongside a particular variant also influence the impact of that variant (e.g., Montgomery & Moore, 2018; Watson & Clark, 2013, 2015). From these factors, the field has begun to assemble a picture of what predicts how a particular utterance will be socially construed by a listener, but there is still much to learn about these factors and how they interact.

CrossMark

This paper examines the contributions of, and interaction between, two factors: a variable's internal constraints and the stylistic (in)congruence between a variant and other covarying cues.

Some work has indicated that a variable's internal constraints, or the linguistic factors that probabilistically predict a particular variant, can affect a variant's social influence (Austen, 2020; Bender, 2005; Drager, 2010; Freitag, 2020; Labov, 2003; Podesva, Reynolds, Callier, & Baptiste, 2015). Other work suggests that listeners' social inferences are affected by the presence of multiple variants together (Austen & Campbell-Kibler, 2022; Campbell-Kibler, 2009; Levon, 2014; Montgomery & Moore, 2018; Pharao, Maegaard, Møller, & Kristiansen, 2014; Watson & Clark, 2013, 2015). In both cases, these effects are likely due to expectations that listeners have built up about the likelihood of encountering a variant in a particular context. Considering English variable (ING), as in *thinking* versus *thinkin'*, from prior experience listeners may have a sense of the relative likelihood of hearing an -in form in a noun (*mornin'*) compared to a verb (*runnin'*), or the likelihood of hearing an -in form in the presence of surrounding speech containing reduced compared to released /t/s, and calibrate their social expectations accordingly.

Using English variable (ING) as a case study, this paper assesses the relative importance to listeners of both variable-internal and multivariable patterns in making social judgments. Two accent rating experiments assess whether listeners' ratings of stimuli are affected not only by the (ING) variant realized but also by (1) the grammatical category of the (ING) word, an example of its internal constraints, and (2) the extent to which the (ING) realization is stylistically congruent with other cues in the stimuli. (2) is assessed by using naturally produced stimuli and spliced versions of the same stimuli, half of the spliced stimuli showing stylistic incongruence between the (ING) realization and the surrounding speech (e.g., an -in spliced in from an utterance originally produced with -ing), and half showing stylistic congruence between the (ING) realization and other cues in the signal (e.g., the -in variant spliced in from an utterance originally produced with -in). The results of these experiments provide a window into the relative weight of within-variable and across-variable patterns in social signaling. They also provide an opportunity to consider the methodological implications of using naturally produced versus spliced stimuli.

The social impact of within-variable patterns: Sensitivity to internal constraints

Both specific instances of sociolinguistic variants and patterns of variant use can carry social meaning (e.g., Bender, 2007; Levon, 2007). That is, an individual realization of a variant can affect listeners' judgments about the speaker, as in the finding that hearing (ING) realized as -in raises the likelihood of perceived Southernness of the speaker as compared with hearing -ing (Campbell-Kibler, 2007). And, individual instances are interpreted in the context of their larger patterning, such that an individual token of -in affects intelligence judgments less when the speaker is already assumed to be Southern (Campbell-Kibler, 2009). Further, the social force of instances and patterns of variant use interact, since the social meaning ascribed to a specific instance of a variant in fact arises from listeners' awareness of its social and linguistic patterning in the first place: variants gain their social meaning in

part because of their presence within socially situated clusters or lects (Bender, 2007; Eckert, 2002; Johnstone, 2016; Ochs, 1992).

A speaker's rate of use of a variant affects listeners' social judgments (e.g., Freitag, 2020; Labov et al., 2011; Levon & Buchstaller, 2015; Levon & Fox, 2014; Wagner & Hesson, 2014), as shown by experiments where listeners hear the same speech sample multiple times with differing rates of a nonstandard variant in each repetition (typically, these differing rates being produced by a splicing procedure) and make a social evaluation of the speaker—often their degree of professionalism—after each sample. Findings indicate that, for variables of sufficient salience like (ING) in American English (Labov et al., 2011; Wagner & Hesson, 2014), hearing higher proportions of the nonstandard variant lowers the social ratings given to the speaker. Labov et al. (2011) explain these findings through the construct of the *sociolinguistic monitor*, proposed to be a cognitive mechanism responsible for tracking sociolinguistic variation for purposes of social evaluation.

Social evaluations can be sensitive to how expected a variant of a single variable is in a given context. That is, in addition to monitoring its frequency overall, listeners can also track where a form tends to occur most often, such as a variant's usage with respect to its internal constraints. For example, Podesva et al. (2015) found stronger social evaluations of /t/-releases in contexts where released /t/ occurs less frequently (word-medially) than in contexts where it is more likely to occur (word-finally). And, Labov (2003; described in Preston, 2011), examined listeners' sensitivity to the grammatical conditioning of variable (ING), of direct relevance to the present study. His sample of ten listeners reported that hearing a passage where -in overwhelmingly occurred in nouns (a grammatical environment more marked for -in) sounded more unnatural than the same passage where -in overwhelmingly occurred in verbs (following typical grammatical constraints) (see also Vaughn, 2022). Bender's (2005) study, perhaps the most developed exploration of this idea to date, showed that the grammatical environment of an instance of copula absence in African American English (AAE) affected the strength of social evaluation, such that copula absence in less attested (more marked) environments increased listeners' negative evaluations of speakers, heightening the social effect of deploying the variant. This effect was confined to those listeners who were most familiar with the variety (see also Austen, 2020). Listeners have also been shown to track the internal constraints of a variable in a linguistic processing task (Vaughn & Kendall, 2018). When listening to sentences containing (ING) words, listeners were faster to classify the variant as -in when it occurred in an (ING) word whose grammatical category strongly favors -ing (noun-like categories) than in those that less strongly favor -ing (verb-like categories). In other words, listeners classified variants faster in grammatical categories for which they have strong expectations about realization when those expectations were violated (i.e., -in realizations in noun-like categories).

Still, a better understanding of the importance of internal constraints to listeners is needed. Internal constraints are central in production studies on the transmission, acquisition, and diffusion of variation. For example, comparing the hierarchy of internal constraints across communities' patterns of production is a foundation of comparative sociolinguistics (e.g., Tagliamonte, 2013). Given the central importance of internal constraints to variationist theory (MacKenzie, 2019; Meyerhoff & Walker,

2007), it is surprising that there has been relatively little attention paid to the role of these constraints in perception, social and otherwise.

The social impact of across-variable patterns: Sensitivity to stylistic congruence

In natural speech settings, listeners encounter streams of linguistic forms, rather than hearing only one at a time. It is in the manipulation of multiple variables together that speakers enact social identities, stances, styles, and personae (e.g., Eckert, 2002; Johnstone, 2016; Levon, 2007, 2014; Podesva, 2008; Zhang, 2005). Speakers' stylistic use of clusters of features means it is likely that listeners have experience with such clusters, and use cues across multiple variables when construing social information. First, listeners make social judgments based on multiple features together, as shown in studies that ask listeners to make social judgments as they are listening to a speech sample, updating their responses over time, allowing analysts to infer the aspects of the speech that affected listeners' inferences about the speaker (e.g., Austen & Campbell-Kibler, 2022; Montgomery & Moore, 2018; Watson & Clark, 2013, 2015). Second, stylistically congruent variants are linked in mental representation (Campbell-Kibler, 2012; Levon, 2007; Vaughn & Kendall, 2019; Wade, 2022). For example, in a novel word game paradigm, American participants produced more Southern-like productions of /aɪ/ (i.e., more monophthongal) after hearing a model talker who never produced tokens of /aɪ/ but whose vowels were otherwise Southern-shifted (Wade, 2022). And, Levon found that pitch range and sibilant duration operated jointly to affect the degree of masculinity British listeners ascribed to speakers, supporting "a gestalt-like understanding of indexicality… whereby linguistic features are not only salient on their own but can also work in clusters to achieve social-indexical significance" (2007:546). In addition, listeners typically encounter stylistically congruent cues together in the signal and expect stylistic congruence among features, being surprised when the covariation among features is mismatched (Vaughn & Kendall, 2018, 2019).

The present study: Within- and across-variable patterns

The present study examines the social force of a form based on both internal constraints and stylistic congruence. The accent rating task is used as a holistic way to elicit listeners' impressions, following work by Campbell-Kibler (2007, 2021) for variable (ING) as well as work in second language speech perception (e.g., Munro & Derwing, 1995). The underspecified nature of "accent" affords listeners a general dimension along which to rate speakers. Although of course the degree of perceived "accent" is always relative to the language variety and experience of the listener, and in fact there are no "unaccented" speakers, the term is adopted here because it is readily used by naive listeners. The crucial measure of interest is the relative accent rating of a stimulus produced with -in as compared to -ing, and whether that difference is conditioned by (1) internal constraints and (2) stylistic (in)congruence.

To assess question (1), listeners' ratings of -in versus -ing stimuli are compared across (ING) words of different grammatical categories. It is expected that accent ratings for -ing forms will be low across grammatical categories, as it is the canonical

form. The measure of interest, then, is whether accent ratings of -in realizations for (ING) words in grammatical categories more marked for -in production (noun-like forms) are greater than -in ratings for grammatical categories where -in is less marked (verb-like forms). In other words, is the social effect of a form's realization modulated by its internal constraints?

To assess (2), Experiment 1 uses naturally produced stimuli, and Experiment 2 uses spliced versions of the same stimuli. In the natural stimuli, speakers produced the stimuli once with (ING) words realized as -ing, and once as -in, and were instructed not to change anything beyond the (ING) realization. Following much prior work (Campbell-Kibler, 2007, inter alia; Labov et al., 2011; Levon & Buchstaller, 2015; Levon & Fox, 2014; Podesva et al., 2015), spliced stimuli were created using a cutting-and-pasting variation of the matched guise methodology, where the -ing realization from the natural -ing "frame" is replaced by an -in realization spliced in from the natural -in production of the sentence (the -in "frame"). In another version, the -ing realization is replaced by the same -ing pasted back into the frame, to maintain any artifacts created by the splicing process. In this study, the same process is also done for the other frame, the natural -in frame sentences, creating four versions of each stimulus.

In general, the splicing method is used to ensure that the (ING) realization itself is the only aspect of the signal that changes across guises. For the purposes of this study, the spliced stimuli offer an additional useful property: half of the stimuli show stylistic congruence between the realization and the frame in which it was produced, and half do not. Thus, comparing participants' behavior when listening to spliced versus natural stimuli and congruent versus incongruent spliced stimuli allows an exploration of whether any reliance on (ING)'s internal constraints in Experiment 1 (when the signal contains only cues stylistically congruent with (ING)'s realization) will still be evident for spliced stimuli (where half of the stimuli are stylistically incongruent between the (ING) variant and the surrounding signal).

Since stretches of natural speech tend to cohere stylistically, there is little chance to test whether listeners are sensitive to the stylistic context of speech patterns. Intentionally introducing incongruence through spliced stimuli here allows for this test; it is not done because of an inherent interest in how listeners respond to artificially created stimuli but because of what those situations can say about what listeners must be doing during the course of regular speech processing. If listeners show different behavior when the relationship between the (ING) variant and the stylistic congruence of the surrounding speech is mismatched, this suggests that listeners have expectations about stylistic congruence. Alternatively, if listeners' behavior does not change when faced with mismatched stimuli, this suggests that the stylistic congruence that occurs in natural speech is not critical to their accent ratings.

In addition to shedding light on these theoretical issues, directly comparing naturally produced versus spliced stimuli in this study has important methodological implications for the field. Splicing has become a dominant approach in matched guise studies, but the consequences of its use have not been fully explored. In an early use of this approach, Campbell-Kibler suggested a need to consider the ramifications of splicing: "Although [the splicing procedure] did not make [the stimuli] strange or

unnatural, it does raise interesting questions regarding exactly what we consider to be 'matched' in matched guise work" (2009:139). Further, most such studies splice realizations of a variable into only one frame, and, in doing so, have conflated the effect of the realization of a variable with the effect of its stylistic (in)congruence with the surrounding speech. Splicing a variant into a frame originally produced in the context of another variant necessarily introduces mismatches between the variant's realization and the style of the surrounding speech. The present design, which creates both stylistically congruent and incongruent versions of stimuli with both realizations of (ING), can disentangle the two and can more generally speak to the consequences of the now common methodological decision to use spliced stimuli.

Experiment 1: Natural stimuli
Methods

Stimuli. The stimuli used in this study were ninety-six sentences (plus four practice sentences), each containing one (ING) word, with a total of forty-six distinct word types, originally used in Vaughn and Kendall (2018, Experiment 3; see Appendix A of that paper for full list of stimuli and details). (ING) words from the following grammatical categories were included: progressive verbs ($n = 48$), gerunds ($n = 11$), adjectives ($n = 16$), nouns ($n = 9$), and two types of pronouns: the two-syllable pronouns *something* and *nothing* (pronoun-2s, $n = 6$), and the three-syllable pronouns *anything* and *everything* (pronoun-3s, $n = 6$); stimulus distribution across grammatical categories was guided by Hazen (2008). Following Vaughn and Kendall (2018), pronoun-3s, adjectives, and nouns are coded here as noun-like categories, which tend to be produced with a lower rate of -in than verb-like categories (here, the category progressive verbs), and gerunds tend to be intermediate. Pronouns are noun-like in their grammar; pronoun-3s are rarely realized with -in, but pronoun-2s are much more commonly realized as -in, and thus pronoun-2s are not classed together with noun- or verb-like categories.

Four female native English speakers produced the stimuli. Three identified as white and one identified as mixed-race; two were from Southern California (ages eighteen and twenty-three), and two were from Oregon (both aged eighteen). Stimuli were recorded in a sound attenuated booth using a Shure SM93 microphone into a Marantz PMD-661 recorder. All speakers were aware of, and could produce, both -ing and -in forms. Speakers were asked to produce the sentences as naturally as possible. Stimuli were amplitude normalized to yield an approximately equal volume across sentences.

Participants. Participants in Experiment 1 were thirty-nine undergraduates from the University of Oregon's Psychology and Linguistics Subject Pool who received partial course credit for their time. Three additional participants were run but excluded because of experimenter or software errors ($n = 2$) or uncorrected hearing loss ($n = 1$). All participants were highly familiar with American English, having learned English at age six or younger. The average age of participants was nineteen years old (min = 18, max = 25). Eighteen participants self-identified as male, twenty as female,

and one as non-binary. Twenty-six participants self-identified as white, one as Black, three as Hispanic, four as Asian, four as mixed-race, and one as other (without identifying more specifically).

Procedure. Each participant heard all ninety-six sentences, divided equally across the four talkers (twenty-four sentences per talker), with sentences randomized per participant. For each sentence, half of participants heard the -ing and half heard the -in version. This counterbalancing was done within talker, such that listeners heard twelve -ing and twelve -in sentences from each talker. Thus, for every listener, each talker's overall rate of (ING) realization, and the rate of (ING) realization overall, was 50% -ing, 50% -in.

Participants completed the task individually seated in a sound-attenuated room in front of a Mac Mini computer, wearing Sennheiser HD-202 headphones. The task was presented using *PsychoPy* (Peirce, 2007). Participants were told they would be listening to English speech and were asked to rate the level of accent in each stimulus on a scale from 1-9, where 1 = no accent and 9 = very strong accent, by pressing the numbers one through nine on the keyboard. Participants were told that all speakers were native English speakers and that they should rate the level of accent of each sentence in relation to the other sentences they hear. Participants could hear each sentence only once. Following the accent rating task, all participants completed a demographic and language background questionnaire.

Results

A total of 3,744 responses were collected (96 items × 39 participants). Before analysis, potentially spurious responses were trimmed by first transforming response reaction times to natural log values and then removing responses greater than ±2.5 standard deviations from each participant's mean log RT from the dataset ($n = 63$).

Figure 1 presents listeners' ratings of each sentence according to realization of the (ING) word, replicating expectations from prior work that stimuli with -in realizations were rated as sounding more accented ($M = 4.18$) than those with -ing realizations ($M = 1.93$), and Table 1 provides numbers of observations and means for factors of interest. Figure 2 plots accent ratings by realization in interaction with grammatical category. Grammatical categories are presented along the x-axis according to the likelihood that a category would be produced with an -in realization in production (following Vaughn & Kendall, 2018), with categories on the left being less likely to be produced as -in and those on the right being more likely. Examining Figure 2, ratings of stimuli produced with -ing did not vary much according to grammatical category, while stimuli realized as -in appear to show higher accent ratings toward the left side of the plot, in grammatical categories where -in is less common in production.

Turning to statistical analysis to assess these observations, accent ratings were modeled using mixed-effect linear regression (with the *lmerTest* package in *R*; Kuznetsova, Brockhoff, & Christensen, 2014), with individual accent rating responses, centered and scaled, as the dependent variable. Models considered random intercepts for participant, stimulus sentence, speaker, and (ING) word, but (ING) word was determined not to improve model fit, and the final model included only random effects for
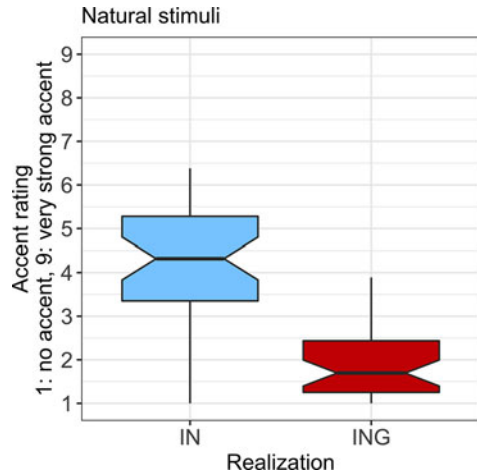
Figure 1. Accent rating for Experiment 1 by realization.

Table 1. Experiment 1 (Natural stimuli) raw observations and mean ratings per factor

		N	Mean
Grammatical Category	Pronoun-3 (ref)	231	3.273
	Adjective	607	3.213
	Noun	343	3.122
	Gerund	423	3.277
	Progressive	1845	2.962
	Pronoun-2	232	2.625
Realization	IN (ref)	1838	4.187
	ING	1843	1.921

participant, sentence, and speaker. The model included fixed effects of realization of (ING) word (dummy coded, -in as reference level), grammatical category of (ING) word (dummy coded, pronoun-3 as reference level), as well as their interaction.[1] An analysis of deviance table (using the *car* package in *R*; Fox & Weisberg, 2011) is shown in Table 2, and the fixed effects are summarized in Table 3.

Statistical modeling confirmed that -in realizations were given significantly higher accent ratings than -ing versions of the same sentences ($\chi^2 = 2341.54$, $p < .001$),[2] as shown in Figure 1, and that this was moderated by a significant interaction with grammatical category ($\chi^2 = 22.67$, $p < .001$), as shown in Figure 2. Grammatical categories more commonly realized with -in in production (progressive verbs and pronoun-2s) showed a smaller difference between ratings of -in and -ing as compared to the reference level pronoun-3s, a category rarely realized as -in. The effect of realization for the noun-like categories, adjectives, and nouns, and gerunds (the other grammatical categories where -in is less common in production), did not significantly differ from the effect of realization for pronoun-3s.
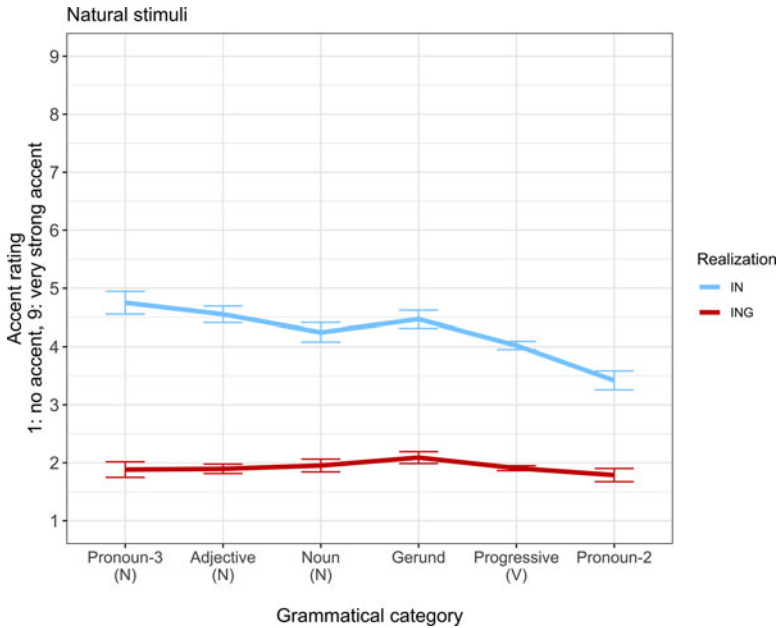
Figure 2. Accent rating for Experiment 1 by realization and grammatical category.

Table 2. Experiment 1 (Natural stimuli) analysis of deviance

	χ^2	DF	Pr(>χ^2)	
Realization	2341.54	1	< 2.2e-16	***
Grammatical Category	10.70	5	0.058	.
Realization:GramCat	22.67	5	0.0004	***

Signif. codes: 0 = ***, 0.001 = **, 0.01 = *, 0.05 = ., 0.1

Discussion

Experiment 1, using natural stimuli, found that listeners rated sentences with (ING) words realized as -in as more accented than when realized as -ing, as expected from prior findings (e.g., Campbell-Kibler, 2007). Further, accent ratings of -in were affected by the grammatical category of the (ING) word, in line with prior work suggesting that internal patterns of variant use can affect social evaluations (e.g., Bender, 2005). When the (ING) word was realized as -ing, accent ratings did not differ across grammatical categories, in keeping with -ing's status as the unmarked form. When realized as -in, however, accent ratings were higher when the -in occurred in grammatical categories where it was less expected from production norms (e.g., noun-like forms), as compared with categories where -in was less marked (e.g., verbs and pronoun-2s). The markedness of the marked variant conditioned its social impact. These results bolster evidence that listeners track the internal constraints of variables and can make use of them when giving accent ratings.

Table 3. Experiment 1 (Natural stimuli) fixed effects

	Estimate	Std. Error	Pr(>|t|)	
(Intercept)	.649	.192	0.009	**
Realization ING	−1.207	.090	< 2e-16	***
Grammatical Category (Adjective)	.056	.107	0.602	
Grammatical Category (Noun)	−.079	.118	0.505	
Grammatical Category (Gerund)	−.017	.113	0.882	
Grammatical Category (Progressive)	−.196	.097	0.046	*
Grammatical Category (Pronoun-2)	−.351	.131	0.008	**
GramCatAdjective:RealizING	.006	.104	0.953	
GramCatNoun:RealizING	.104	.116	0.371	
GramCatGerund:RealizING	.083	.111	0.456	
GramCatProgressive:RealizING	.221	.095	0.020	*
GramCatPronoun-2:RealizING	.371	.128	0.004	**

Signif. codes: 0 = ***, 0.001 = **, 0.01 = *, 0.05 = ., 0.1

Experiment 2: Spliced stimuli

Experiment 1 confirmed that listeners can use variable-internal patterns when making social judgments about speech, finding an interaction between grammatical category and (ING) realization when the stimuli were naturally produced and, therefore, all covarying cues were congruent with the realization of (ING). Vaughn and Kendall (2018), with the same stimuli, found different behavior regarding grammatical category information in natural versus spliced stimuli: when classifying whether they heard -in or -ing, listeners used grammatical category expectations when listening to spliced stimuli (Experiment 2), but not naturally produced stimuli (Experiment 3). In different tasks, however, listeners may weigh their use of grammatical expectations versus stylistic congruence in different manners.

The effect of stylistic congruence on listeners' use of (ING)'s realization and its internal constraints might surface in several ways in the accent rating task. It may be that the lack of congruence in some spliced stimuli is the most salient factor to listeners, making sentences with incongruent cues (where frames and realizations mismatch) rated as more accented than those with congruent cues (where frames and realizations match), irrespective of the realization of (ING) or grammatical category. Or it may be that stylistic congruence interacts with the realization of (ING), where (ING) realization predicts accent rating but the realization's ability to signal social meaning may be mitigated in cases when stylistic cues do not match the realization. Finally, stylistic congruence may interact with the grammatical category effect: Given that many cues covary with the realization of (ING), the grammatical category effect may get swamped by the information carried across other cues in the spliced stimuli. In this case, the grammatical category pattern evident in -in realizations in Experiment 1 may not be present at all in Experiment 2, or may only be present for stylistically congruent (matching) stimuli.

Types of stylistic covariation present in these stimuli are illustrated in Vaughn and Kendall (2019). In that paper, acoustic analyses were conducted on the natural stimuli produced by these four talkers; the ninety-six stimuli used in Experiment 1 are a subset of the stimuli analyzed there. Five phonetic features were selected for analysis because they also index aspects of -in's indexical field, namely Southernness and/or casualness: /aɪ/-glide length, spectral proximity of the mid-front vowels /e/ and /ɛ/, duration of the lax vowels /ɛ/ and /ɪ/, speaking rate, and prevalence of release and reduction of intervocalic /t/. Statistical models determined whether the production of these features differed significantly across sentences produced with -ing versus -in realizations. Findings indicated that, compared to the -ing versions of the same sentences, the -in versions had significantly shorter /aɪ/-glides, more proximate mid-front vowels /e/ and /ɛ/, longer lax vowels /ɛ/ and /ɪ/, and more reduced /t/ productions. Speaking rate did not pattern according to (ING) realization, instead showing different patterns by speaker across -ing and -in. Thus for four of the five features, results indicated that speakers' productions covaried with (ING) realization in ways that were stylistically congruent with -in's social meanings of casual (e.g., more reduced /t/ productions) or Southern (e.g., shorter /aɪ/-glides, despite the four speakers being from the Western US). It is thus evident that the realization of (ING) is not the only factor that varies across the naturally produced -ing and -in frames (and therefore was not the only factor driving the accent ratings in Experiment 1). Because of these patterns of covariation in the natural stimuli, splicing (ING) realizations across frames for Experiment 2 will necessarily introduce stylistic mismatches between (ING) realization and other features. For example, when splicing -ing into an -in frame, cues in -in frames that index casualness will be incongruent with -ing's more formal status.

Methods

Stimuli

The stimuli used in this study are spliced versions of those in Experiment 1 (from Vaughn and Kendall, 2018, Experiment 2). For the -ing version and the -in version (or "frame") of each sentence, splicing was done by identifying boundaries of each (ING) realization using auditory and spectrographic cues in *Praat* (Boersma & Weenink, 2019) and selecting the nearest zero-crossing in the waveform. Then, the realizations from the opposite version of each sentence (e.g., -in) were pasted into the other version (e.g., -ing), replacing the original realization for each frame, to create *mismatching* or *stylistically incongruent* stimuli (e.g., -in realization in -ing frame). To create *matching* or *stylistically congruent* stimuli (e.g., -ing realization in -ing frame), the realization was copied and pasted back into the same sentence frame opened in a different window, making it unlikely that identical zero-crossings were selected. Thus, any artifacts of the splicing procedure itself should occur in both matched and mismatched stimuli. All stimuli were RMS amplitude normalized to yield an approximately equal volume across stimuli.

Participants

Due to COVID-19, participants in this study were 117 participants recruited on Amazon's Mechanical Turk rather than in the lab. Because of the design of the

spliced stimuli, more participants were required than in Experiment 1 to ensure that an adequate number of participants heard each of the four versions of each stimulus. Mechanical Turk participants were restricted to IP addresses in the United States and were paid the equivalent of $10/hour for their participation. Twenty-seven additional participants were run but were excluded because they learned English after age 6 ($n = 2$), or because they failed attention checks, indicating that they were not adequately completing the task ($n = 25$). The average age of participants was thirty-eight (min = 20, max = 71). Seventy-six participants self-identified as male, thirty-nine as female, one as nonbinary, and one preferring not to report. Ninety-one participants self-identified as white, seven as Black, five as Hispanic, four as Asian, eight as mixed-race, one as other, and one preferring not to report.

Procedure

The procedure and instructions to participants were identical to those in Experiment 1, with only the presentation software and experimental stimuli differing. The task was presented in a web browser using *FindingFive* (FindingFive Team, 2019), and participants were asked to complete the task using headphones. As in Experiment 1, each participant heard all ninety-six sentences, divided equally across the four talkers, with sentences randomized per participant. For each sentence, half of the participants heard an -ing and half heard an -in realization, and within each of those realizations, half heard the version where frame and realization matched, and half heard the mismatched version. In this way, an equal number of participants heard all four versions of each sentence. This counterbalancing was done within talker, such that listeners heard twelve -ing and twelve -in sentences from each talker. Thus, as in Experiment 1, for every listener each talker's overall rate of (ING) realization, and the rate of (ING) realization overall, was 50% -ing, 50% -in.

Results

11,232 responses were collected in total (96 items × 117 participants). Following Experiment 1, log RTs greater than ±2.5 standard deviations from each participant's mean log RT were trimmed ($n = 204$). Table 4 provides numbers of observations and means for factors of interest. Looking first at the raw data plotted in Figure 3 to consider the effect of realization on accent ratings, the effect appears as expected from prior work and Experiment 1: accent ratings were higher for stimuli with -in realizations ($M = 4.02$) than with -ing realizations ($M = 2.76$). Further, this pattern appears to hold for stimuli where the frame and realization matched and also where they mismatched. However, stylistic congruence between frame and realization seems to amplify the effect of realization, with -in realizations from -in frames having the highest accent ratings ($M = 4.41$), -ing realizations from -ing frames having the lowest ($M = 2.35$), and mismatched stimuli falling in between (-ing realization/-in frame $M = 3.16$, -in realization/-ing frame $M = 3.63$).

Figure 4 displays grammatical category in interaction with realization. In stimuli where frame and realization matched (left panel of Figure 4), the pattern appears similar to Experiment 1: accent ratings of -in realizations (with -in frames) seem to vary systematically by grammatical category (with -in realizations for noun-like categories

Table 4. Experiment 2 (Spliced stimuli) raw observations and mean ratings per factor

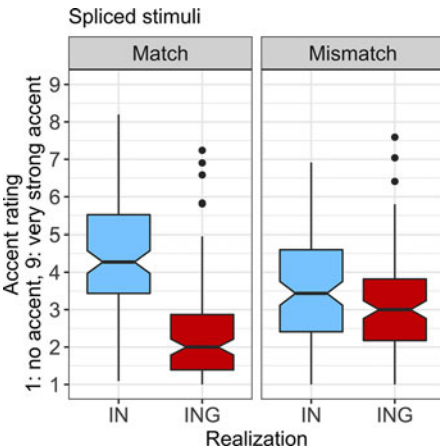|		N	Mean
Frame.Realization	Match (ref)	5511	3.381
	Mismatch	5517	3.396
Grammatical Category	Pronoun-3 (ref)	686	3.512
	Adjective	1846	3.608
	Noun	1033	3.421
	Gerund	1259	3.524
	Progressive	5516	3.321
	Pronoun-2	688	2.917
Realization	IN (ref)	5520	4.018
	ING	5508	2.757

Figure 3. Accent rating for Experiment 2 by realization and whether stimulus realization and frame matched or mismatched.

receiving higher accent ratings than verb-like categories), while accent ratings of -ing realizations (with -ing frames) are less affected by grammatical category. However, in stimuli where frames and realizations mismatched (right panel in Figure 4), the same is not true; here, both -ing and -in realizations are variable by grammatical category, and systematicity is less apparent.

Turning to statistical modeling, the same model as in Experiment 1 was used, adding in the new factor of match or mismatch between frame and realization (dummy coded with match as the reference level). Because of the *a priori* interest in the effect of frame and realization match on (ING) realization, grammatical category, and their interaction, the three-way interaction and all two-way interactions of these factors were included in the model. Results are presented in an analysis of deviance table (Table 5) and the fixed effects model output (Table 6).

The most relevant findings are highlighted here, with full model results given in the tables. First, as in Experiment 1, the main effect of realization affected accent
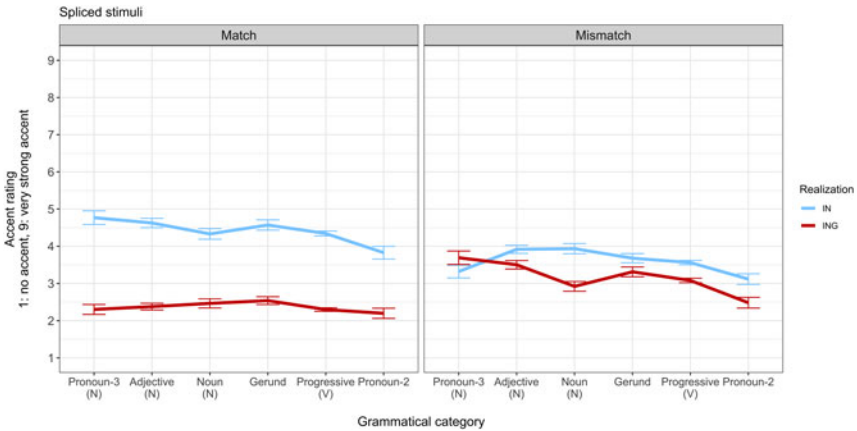
Figure 4. Accent ratings for Experiment 2 by grammatical category, realization, and match/mismatch between frame and realization.

Table 5. Experiment 2 (Spliced stimuli) analysis of deviance

	χ^2	DF	PR($>\chi^2$)	
Frame.Realization	0.041	1	.840	
Grammatical Category	8.700	5	.122	
Realization	1706.080	1	< 2e-16	***
Frame.Realization:Grammatical Category	7.721	5	.263	
Frame.Realization:Realization	689.157	1	< 2e-16	***
Grammatical category:Realization	17.220	5	.005	**
Frame.Realization:Grammatical Category:Realization	45.956	5	9.27e-09	***

Signif. codes: 0 = ***, 0.001 = **, 0.01 = *, 0.05 = ., 0.1

ratings ($\chi^2 = 1710.12$, $p < .0001$), with -in realizations given higher accent ratings than -ing. Importantly, this main effect was moderated by realization's significant interaction with frame-realization match ($\chi^2 = 689.16$, $p < .0001$), grammatical category ($\chi^2 = 17.22$, $p = .004$),[3] and the three-way interaction among these factors ($\chi^2 = 45.96$, $p < .0001$). The significant interaction between realization and frame-realization match confirms the pattern seen in Figure 3, showing that the difference in accent ratings between -in and -ing realizations was larger for stimuli with frame-realization match rather than mismatch. Interestingly, there was not a significant main effect of frame-realization match:[4] it was not the case that mismatches were rated as more accented than matches, regardless of realization. Rather, frame-realization moderated the effect of realization.

The significant three-way interaction between frame-realization match, grammatical category, and realization confirms the pattern evident in Figure 4: the difference in accent ratings between realizations across grammatical categories was larger when

Table 6. Experiment 2 (Spliced stimuli) fixed effects

	Estimate	Std. Error	Pr(>|t|)	
(Intercept)	.462	.192	0.041	*
Frame.Realization Mismatch	−.575	.076	3.38e-14	***
Grammatical Category (Adjective)	.093	.115	0.418	
Grammatical Category (Noun)	−.037	.126	0.768	
Grammatical Category (Gerund)	.002	.122	0.987	
Grammatical Category (Progressive)	−.048	.104	0.643	
Grammatical Category (Pronoun-2)	−.181	.140	0.199	
Realization ING	−.962	.075	< 2e-16	***
Frame.RealizMismatch:GramCatAdjective	.257	.088	0.004	**
Frame.RealizMismatch:GramCatNoun	.394	.097	4.80e-05	***
Frame.RealizMismatch:GramCatGerund	.195	.094	0.038	*
Frame.RealizMismatch:GramCatProgressive	.229	.080	0.004	**
Frame.RealizMismatch:GramCatPronoun-2	.271	.106	0.010	*
Frame.RealizMismatch:RealizING	1.190	.106	< 2e-16	***
GramCatAdjective:RealizING	.0328	.087	0.705	
GramCatNoun:RealizING	.125	.097	0.196	
GramCatGerund:RealizING	.081	.094	0.387	
GramCatProgressive:RealizING	.081	.080	0.307	
GramCatPronoun-2:RealizING	.211	.107	0.045	*
Frame.RealizMismatch:GramCatAdjective:RealizING	−.427	.123	0.0005	***
Frame.RealizMismatch:GramCatNoun:RealizING	−.822	.136	1.51e-09	***
Frame.RealizMismatch:GramCatGerund:RealizING	−.452	.135	0.0008	***
Frame.RealizMismatch:GramCatProgressive: RealizING	−.507	.113	6.54e-06	***
Frame.RealizMismatch:GramCatPronoun-2: RealizING	−.745	.150	7.49e-07	***

Signif. codes: 0 = ***, 0.001 = **, 0.01 = *, 0.05 = ., 0.1

frames and realizations matched than when they mismatched. Listeners appeared to use grammatical category information in assigning accent ratings more when frames and realization were stylistically congruent. The significant two-way interaction between grammatical category and realization confirms this pattern for matched trials (displayed in Table 6), though the only comparison that reached significance there was the difference in realization between pronoun-3 and pronoun-2, the grammatical categories with more extreme expectations from production. When frames and realizations mismatched, the realization by grammatical category interaction was mitigated.

Discussion

In Experiment 2, with spliced stimuli, the expected effect of realization on accent ratings was replicated, with -in realizations assigned higher accent ratings than -ing realizations. The stylistic congruence between frame and realization moderated the effect of realization, where congruent stimuli showed more extreme accent ratings based on realization than did incongruent stimuli. In other words, incongruence did not completely swamp the role of realization, but instead both (ING) realization and the stylistic context in which the (ING) realization occurred affected listeners' ratings.

The grammatical category by realization interaction was also conditioned by the congruence between frame and realization. For stylistically incongruent stimuli, there was a disruption of the role of grammatical category in affecting the difference between accent ratings of -in and -ing realizations; in mismatched stimuli, listeners' expectations about -in's degree of markedness played less of a systematic role in their accent ratings compared to the naturally produced stimuli in Experiment 1 and the stylistically congruent (matched) stimuli in Experiment 2. This pattern suggests that the mismatching cues in those stimuli had more of an impact on accent ratings than did (ING)'s grammatical category information. Stylistic incongruence appears to have overshadowed any potential usefulness of listeners' knowledge of internal conditioning information, though notably it did not overshadow the usefulness of (ING) realization.

General discussion

The findings reported here support a role for both variable-internal and across-variable patterns, and their interaction, in social signaling. Results confirm that listeners have knowledge of conditioning constraints, supporting prior work on (ING) and other variables (e.g., Bender, 2005; Vaughn & Kendall, 2018). Further, the grammatical category conditioning of (ING), a constraint also used in linguistic processing (in Vaughn & Kendall, 2018), is available for use in accent rating, a task that is more tied to social evaluation. And, the stylistic frame in which the realization of (ING) occurred also affected accent ratings: although listeners made use of (ING)'s realization when assigning accent ratings in both naturally produced and spliced stimuli, the effect of realization was diminished for stylistically incongruent stimuli. Finally, the grammatical constraints governing (ING) were not used by listeners in assigning ratings when the stimuli were stylistically incongruent. The following sections discuss some implications of these findings.

Role of markedness/surprisal in social signaling

One interpretation of the finding that listeners use probabilistic information about (ING)'s internal constraints in the accent rating task is that a form's markedness, or surprisal, in a given context is a part of what predicts the strength of its social signaling (Bender, 2007; Jaeger & Weatherholtz, 2016; Rácz, 2013). That is, the deployment of a marked variant in an unexpected context amplifies its social effect.

Speakers' grasp of the constraints on sociolinguistic variation is likely part of what allows them to construct styles in inventive ways in the first place; being able to anticipate their interlocutor's degree of surprise at the deployment of a particular form—in

concert with other forms—is part of the process of identity construction in language. Speakers' knowledge about the linguistic and social factors that condition a form's use in a particular context lets them predict whether their use of a form may amplify, or even block or subvert, an expected social meaning for a listener. It may be that the same expectation-based mechanism that produces the surprisal effect in this study also results in "blocking" or "indexical bulletproofing" in cases where a form occurs in a rare context (see, for example, Campbell-Kibler, 2011; Levon, 2007, 2014; Pharao et al., 2014). Going forward, the separate literatures documenting markedness effects on social signaling from *social* conditioning (e.g., Campbell-Kibler, 2011; Levon, 2007, 2014; Pharao et al., 2014; Stecker, 2020) and from *linguistic* conditioning (e.g., Bender, 2005; Podesva et al., 2015) would benefit from being considered together, contributing to a larger account of the ways that markedness can drive social indexing.

Converging evidence from other areas points to a role of markedness in social processing. Results from recent artificial language learning experiments suggest that sociolinguistic variants that are more unexpected given particular social contexts are more learnable (Lai, Rácz, & Roberts, 2020), as are variants that are associated with more salient social cues (Rácz, Hay, & Pierrehumbert, 2017; Sneller & Roberts, 2018). The impact of a form's typicality in context is also predicted by theories of spoken word processing that assign different attentional weights to forms with different degrees of social significance (Sumner, Kim, King, & McGowan, 2014).[5]

There are many more open questions regarding how markedness/surprisal affects social perception; two examples are discussed here. First, the extent to which a particular variable-internal pattern is made use of by listeners likely depends on the variable and the patterning of that variable within and across communities. For example, in this study testing variable (ING) listeners participated from across demographic categories and dialect boundaries since the grammatical conditioning of (ING) has been shown to be stable across a variety of Englishes and does not appear to often interact with external constraints. Thus, it was expected that a variety of listeners would be aware of and able to use the markedness of (ING)'s internal constraints for social judgments. However, understanding how listeners build up representations about the markedness of internal constraints upon encountering changing or unfamiliar variants or speech communities is an important open question (see Bender, 2005; Levon & Fox, 2014). For example, the extent to which an internal constraint is informative about social factors is likely to modulate this effect (e.g., Villareal, Clark, Hay, & Watson, 2021; Wolfram, Childs, & Torbert, 2000). Second, there are open questions regarding the level at which within-variable patterns are represented in the mind. For example, does the grammatical category in which the variant is used best explain how markedness patterns are stored, or is there word-specific information that could account for this knowledge as well (see Vaughn, 2022)? There is ample room for future work to explore these issues.

Processing across tasks and types of stimuli

Although this study demonstrates that listeners *can* use grammatical category knowledge about the likelihood of (ING) realizations in their accent ratings of a sentence, it also demonstrates that listeners do not always use such knowledge. Judgments based

on spliced stimuli did not reflect this grammatical knowledge (taking together matching and mismatching stimuli). Interestingly, this pattern is the opposite of what Vaughn and Kendall (2018) found with the same stimuli in a different task, where listeners' judgments about which realization of (ING) they heard reflected (ING)'s grammatical constraints only for spliced stimuli. Vaughn and Kendall reasoned that the incongruence of the mismatches in the spliced experiment put listeners in a mode of processing where they could rely less on the bottom-up information in the signal to accurately cue the realization of (ING). With spliced stimuli, listeners relied more on top-down expectations about grammatical category to accomplish that task.

These seemingly opposing findings across studies are readily explained by considering the nature of the tasks. The variant classification task asks listeners to focus specifically on the realization of the (ING) variable itself. In that case, mismatches between frame and realization impede listeners' ability to do this task, leading them to use other information at their disposal: their knowledge about (ING)'s grammatical category conditioning. In contrast, the accent rating task involves holistic judgments of the stimulus, where the realization of (ING) is only one component among the many linguistic features across the sentence. In this task, given that other features (e.g., /aɪ/-glide length, /t/-release, and position and duration of certain vowels) covary in stylistically congruent ways with the realization of (ING) in these stimuli (Vaughn & Kendall, 2019), it is not surprising that the rest of the stimulus contains information that affects accent ratings. Here, the congruence between constellations of cues matters more than the constraints governing just one variable.

The pattern of results across these two studies indicates that the type of stimuli heard affects listeners' use of variable-internal probabilistic conditioning expectations in ways appropriate to their attentional goals. Mental representations of probabilistic conditioning information are apparently available to listeners for use in both linguistic and social processing, and the processing system calls on those representations as needed for the task at hand. This suggests that modular, separate constructs like the sociolinguistic monitor, thought to track and store information about sociolinguistic variation *for use in social evaluation* (Labov et al., 2011), may not be necessary (see also Campbell-Kibler, 2021). Listeners show sensitivity to probabilistic information about what conditions (ING) variation in both social and linguistic tasks, albeit in the manner most appropriate to the task.

Splicing methodology and the importance of stylistic congruence

The results of this study have significant methodological implications. Much recent work has employed the splicing procedure because of a well-motivated interest in better understanding the impact of individual variables. However, these findings present mounting evidence that this methodological decision has consequences: listeners' behavior changes across stylistically congruent versus incongruent stimuli.

Since naturally produced stimuli are of the sort that listeners encounter in actual speech, of course, the current findings suggest that the grammatical category of -in may be readily available for listeners to use in making judgments about speakers in everyday settings, a noteworthy finding. However, given that the splicing procedure is standard practice in the field, it is worth better understanding why the same findings

did not hold for spliced stimuli. It may be that the artifacts introduced in the process of splicing were overly disruptive. Or it may be that the presentation of the stylistically incongruent stimuli (randomized together with the stylistically congruent stimuli) meant that listeners could not consistently rely on the congruence between all the bottom-up information in the stimuli and thus allocated processing resources differently than had all realizations been matched to their frame. Prior work lends support to the latter explanation. For example, listeners have been shown to put less weight on a particular acoustic cue, even one that is relevant for the current stimulus, when they are in a context in which the cue is less reliable (McQueen & Huettig, 2012).

Listeners' sensitivity to subtle phonetic incongruence has been well-documented in other domains (see Sumner et al., 2014 for a review). Sumner (2013) found that two variants of words with medial /nt/ sequences (the more careful *splin[t]er* and more casual *splin[_]er* for the word "splinter") could prime a semantically related word (e.g., "wood"), but only when the variants were congruent with the overall style of the word, not when the more casually articulated n[_] variant was housed in a more carefully articulated frame.

Uncovering the precise features that contributed to the effects of stylistic incongruence in these stimuli, and accounting for the exact pattern of results in the mismatched stimuli, are clear next steps. There are many possible sources of these patterns, from speakers' productions of particular items, to commonalities in productions across grammatical categories or speakers, to listeners' expectations about these productions, to intersections of all of these factors. This paper was not designed to test these possibilities, and as such there are simply not enough stimuli to tease apart patterns in a systematic way. Although Vaughn and Kendall (2019) established that there are systematic acoustic differences between -in and -ing sentences for the linguistic features measured there, it is not possible to test differences on a by-grammatical category basis: those features do not occur equally, or sometimes even occur at all, across all grammatical categories. Future work explicitly designed to systematically vary potential covarying cues could make inroads toward understanding the dynamics of ratings made to stylistically mismatched stimuli. For now, this paper establishes that there is reason to expect that listeners are indeed sensitive to the stylistic incongruence induced by splicing.

Going forward, what might this documented role of stylistic (in)congruence mean for the use of spliced stimuli in matched guise tasks? First, the impact of realization of (ING) appears robust to the stylistic congruence or incongruence between frame and realization. In fact, the present findings (see Figure 3 in particular) suggest that the impact of realization may be underestimated in prior studies using only one frame (as often is the case in sociolinguistic studies using stimuli excised from sociolinguistic interviews, where only one version of the sentences is available), since the cues from the original frame appear to counteract the effect of realization. Of course, the potential for congruent cues to amplify the social impact of the realization, which is what was observed in Experiment 2, is part of the reason why the splicing procedure is conducted in the first place. But as we observe here, the other side of that coin is that splicing in cues from an incongruent frame dampens the effect of realization. This finding underscores the idea that listeners are using more than just the realization of a single variant to make social evaluations.
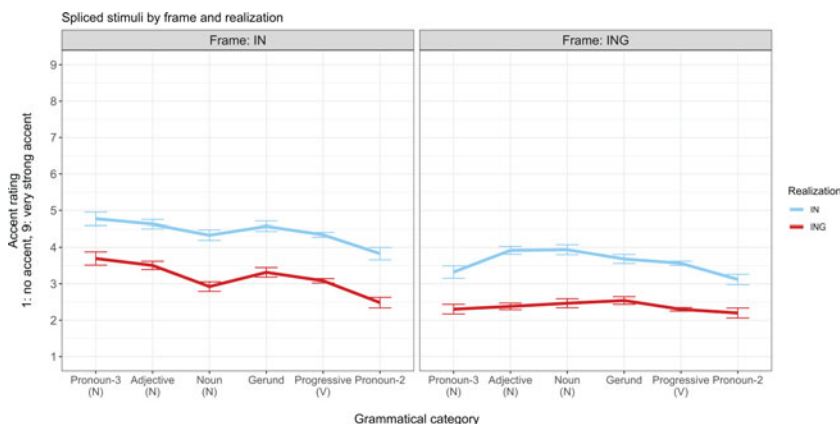
Figure 5. Accent ratings for Experiment 2 by grammatical category, realization, and frame.

Although this paper's goal was to understand the effect of congruence between frame and realization (as shown in Figure 4), the same data can be examined according to the stimuli's original frame, as shown below in Figure 5. Doing so can provide more concrete guidance for future work, as it emulates a methodological decision often made when using spliced stimuli: splicing realizations into only one frame. What is immediately evident from Figure 5 is that if only one frame had been used, which frame was selected would have made a difference. First, the -in frame produced higher accent ratings than the -ing frame across realizations.

Further, there are differences across frames in terms of how grammatical category was reflected in accent ratings. For example, stimuli created by splicing both realizations into the -in frame (Figure 5, left panel) appear to show traces of grammatical category effects even for -ing realizations. This is surprising given that -ing is licensed across grammatical categories. However, rather than undermining the current results, this pattern lends further support to the power of the covarying cues in the frame. That is, speakers' original productions of covarying cues may have been affected by (ING)'s grammatical category. Speculatively, it may be that in aiming to produce -in in a context where it is less licensed grammatically (noun-like forms), speakers heightened their use of other stylistically covarying features in an effort to make the production of -in feel less unnatural. In contrast, the pattern of results for stimuli created by splicing both realizations into the -ing frame (Figure 5, right panel) appears more similar to the findings from the natural or matched stimuli, where listeners' expectations about the markedness of -in across grammatical categories drives the pattern. This general pattern suggests that in interpreting listeners' social evaluations of spliced stimuli, at least for (ING) in read speech, it is important to acknowledge the role of internal constraints not just for listeners, but also stimulus speakers.

One concrete methodological takeaway, then, is that if choosing to splice into only one frame, it is more advisable to use the same frame consistently across all stimuli rather than to select the frame on a stimulus-by-stimulus basis (i.e., for read sentences, selecting which frame sounds "more natural" for each sentence, or, for stimuli excised from

interviews, selecting the frame that happens to be present). With only one frame, the potential effects of that frame can at least be taken into account when interpreting results.

In sum, these observations suggest that sociolinguistic experimentalists carefully consider several factors when determining the best methodology for a study. The research question and dependent and independent variables of interest should shape decisions about the type of stimuli to use. For example, in a given study, is it more important to allow other variables to naturally covary with the variable of interest, or to artificially create instances of stylistic incongruence via splicing? And, if splicing is the preferred method, is it more justified to use naturally produced stimuli (i.e., from a sociolinguistic interview), or to produce new stimuli (see also Tamminga, 2017)? Moreover, if using splicing, should both frames be used in order to observe the effects of realization across both stylistically congruent and incongruent stimuli, or can one frame be selected consistently and factored into how the results are interpreted? Triangulating between several carefully designed tasks and types of stimuli can also be useful.

Conclusion

This paper investigated how several patterns of variation, internal constraints and stylistic congruence, affect how the realization of variable (ING) is socially interpreted by listeners. Results indicate that listeners were sensitive to (ING)'s internal constraints in an accent rating task, but those constraints were only used when the natural stylistic congruence between (ING)'s realization and other cues was not disrupted by using spliced stimuli. Congruence between the (ING) realization and surrounding forms likely enabled listeners to attend to (ING)'s internal constraints. These findings provide converging evidence for two central concepts in variationist sociolinguistic work—a variable's internal constraints and the stylistic covariation among variables —from a methodology other than production. Here, (ING)'s internal grammatical constraints are an example of a pattern that occurs within a variable, and stylistic congruence was operationalized to examine contributions of patterns across variables. Future work could usefully examine how much listeners use other within-variable and across-variable patterns, and their interactions, in making social evaluations.

Notes

1. Dummy coding was used here for maximum comparability to Experiment 2, where it is most useful for ease of interpretation.

2. Dummy coding, though desirable for interpreting the interactions of interest, does complicate the interpretation of main effects. To ensure proper interpretation of all main effects, confirmation is presented from a separate model, identical except for sum coding rather than dummy coding all fixed effects involved in interactions. Confirming the results with sum coded factors ensures that the main effects are generalizable to contexts beyond the reference levels of the factors involved in the interactions. The main effect of realization here was also significant in a sum coded statistical model of the same data.

3. The main effect of realization, and the two significant two-way interactions, were also significant in a model where the three fixed effects were sum coded rather than dummy coded.

4. The main effect of frame-realization match was also not significant in the sum coded model.

5. Still other areas, such as recent work in sociopragmatics, also explore markedness' effect on social signaling (e.g., Acton & Potts, 2014; Beltrama & Staum Casasanto, 2021). Because in that literature markedness tends not to be derived from frequency-based expectations but rather from semantic properties or

lexical characteristics, it is not discussed here, yet a complete account of how markedness affects social processing must incorporate such data.

Competing interests. The author declares none.

References

Acton, Eric. K, & Potts, Christopher. (2014). That straight talk: Sarah Palin and the sociolinguistics of demonstratives. *Journal of Sociolinguistics* 18(1):3–31.

Austen, Martha. (2020). *The Role of Listener Experience in Perception of Conditioned Dialect Variation.* Unpublished doctoral dissertation. The Ohio State University.

Austen, Martha, & Campbell-Kibler, Kathryn. (2022). Real-time speaker evaluation: How useful is it, and what does it measure? *Language* 98(2):e108-130.

Beltrama, Andrea, & Casasanto, Laura S. (2021). The social meaning of semantic properties. In L. Hall-Lew, E. Moore, & R. Podesva (Eds.), *Social meaning and linguistic variation: Theorizing the third wave.* Cambridge: Cambridge University Press.

Bender, Emily M. (2005). On the boundaries of linguistic competence: Matched-guise experiments as evidence of knowledge of grammar. *Lingua* 115(11):1579-98.

Bender, Emily M. (2007). Socially meaningful syntactic variation in sign-based grammar. *English Language and Linguistics*, 11(2):347-81.

Boersma, Paul, & Weenink, David. (2019). Praat: Doing phonetics by computer [Computer program]. Version 6.0.56. http://www.praat.org/.

Campbell-Kibler, Kathryn. (2007). Accent, (ING), and the social logic of listener perceptions. *American Speech* 82:32-64.

Campbell-Kibler, Kathryn. (2009). The nature of sociolinguistic perception. *Language Variation and Change* 21:135-56.

Campbell-Kibler, Kathryn. (2011). Intersecting variables and perceived sexual orientation in men. *American Speech* 86:52-68.

Campbell-Kibler, Kathryn. (2012). The implicit association test and sociolinguistic meaning. *Lingua* 122:753-63.

Campbell-Kibler, Kathryn. (2021). Deliberative control in audiovisual sociolinguistic perception. *Journal of Sociolinguistics* 25(2):253-71.

Drager, Katie. (2010). Sensitivity to grammatical and sociophonetic variability in perception. *Laboratory Phonology* 1(1):93-120.

Eckert, Penelope. (2002). Constructing meaning in sociolinguistic variation. Paper presented at the *Annual Meeting of the American Anthropological Association*, November 20-24, New Orleans, Louisiana.

Eckert, Penelope. (2012). Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41:87-100.

FindingFive Team (2019). FindingFive: A web platform for creating, running, and managing your studies in one place. FindingFive Corporation (nonprofit), New Jersey, USA. https://www.findingfive.com

Fox, John, & Weisberg, Sanford. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage.

Freitag, Raquel Meister Ko. (2020). Effects of the linguistics processing: palatals in Brazilian Portuguese and the sociolinguistic monitor. *University of Pennsylvania Working Papers in Linguistics* 25(2):4.

Hazen, Kirk. (2008). (ING): A vernacular baseline for English in Appalachia. *American Speech* 83:116-40.

Jaeger, T. Florian, & Weatherholtz, Kodi. (2016). What the heck is salience? How predictive language processing contributes to sociolinguistic perception. *Frontiers in Psychology* 7:1115.

Johnstone, Barbara. (2016). Enregisterment: How linguistic items become linked with ways of speaking. *Language and Linguistics Compass* 10(11):632-643.

Kuznetsova, Alexandra, Brockhoff, Per B., & Christensen, Rune H. B. (2014). LmerTest: Tests for random and fixed effects for linear mixed effect models. R package, version 2.0-3.

Labov, William. (2003) "Re-defining 'the same' as a mechanism of linguistic change." Forum Lecture, *Linguistic Society of America Institute*. East Lansing: Michigan State University, August 7.

Labov, William, Ash, Sharon, Ravindranath, Maya, Weldon, Tracey, Baranowski, Maciej, & Nagy, Naomi. (2011). Properties of the sociolinguistic monitor. *Journal of Sociolinguistics* 15(4):431-63.

Lai, Wei, Rácz, Péter, & Roberts, Gareth. (2020). Experience with a linguistic variant affects the acquisition of its sociolinguistic meaning: An alien-language-learning experiment. *Cognitive Science* 44(4):e12832.

Levon, Erez. (2007). Sexuality in Context: Variation and the Sociolinguistic Perception of Identity. *Language in Society* 36(4):533-54.

Levon, Erez. (2014). Categories, stereotypes and the linguistic perception of sexuality. *Language in Society* 43(5):539-66.

Levon, Erez, & Buchstaller, Isabelle. (2015). Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing. *Language Variation and Change* 27:319-48.

Levon, Erez, & Fox, Sue. (2014). Social salience and the sociolinguistic monitor: A case study of ing and th-fronting in Britain. *Journal of English Linguistics* 42(3):185-217.

MacKenzie, Laurel. (2019). Perturbing the community grammar: Individual differences and community-level constraints on sociolinguistic variation. *Glossa: A journal of general linguistics* 4(1).

McQueen, James M., & Huettig, Falk. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America* 131(1):509–17.

Meyerhoff, Miriam, & Walker, J. (2007). The persistence of variation in individual grammars: Copula absence in urban sojourners and their stay-at-home peers, Bequia (St. Vincent and the Grenadines). *Journal of Sociolinguistics* 11(3):346-66.

Montgomery, Chris, & Moore, Emma. (2018). Evaluating S(c)illy voices: The effects of salience, stereotypes, and co-present language variables on real-time reactions to regional speech. *Language* 94(3):629-61.

Munro, Murray J., & Derwing, Tracey M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45(1):73-97.

Ochs, Elinor. (1992). Indexing gender. In A. Duranti and C. Goodwin (Eds.), *Rethinking context: language as an interactive phenomenon*. Cambridge: Cambridge University Press. 335-58.

Peirce, Jonathan W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods* 162(1-2):8-13.

Pharao, Nicolai, Maegaard, Marie, Møller, Janus Spindler, & Kristiansen, Tore. (2014). Indexical Meanings of [s+] Among Copenhagen Youth: Social Perception of a Phonetic Variant in Different Prosodic Contexts. *Language in Society* 43(1):1-31.

Podesva, Robert J. (2008). Three sources of stylistic meaning. *Texas Linguistic Forum* 51:134-43.

Podesva, Robert J., Reynolds, Jermay, Callier, Patrick, & Baptiste, Jessica. (2015). Constraints on the social meaning of released /t/: A production and perception study of U.S. politicians. *Language Variation and Change* 27:59-87.

Preston, Dennis R. (2011). The power of language regard-discrimination, classification, comprehension, and production. *Dialectologia: Revista electrònica*, 9-33.

Rácz, Péter. (2013). *Salience in sociolinguistics: A quantitative approach* (Vol. 84). Berlin: Walter de Gruyter Mouton.

Rácz, Péter, Hay, Jennifer B., & Pierrehumbert, Janet B. (2017). Social salience discriminates learnability of contextual cues in an artificial language. *Frontiers in Psychology* 8(51).

Sneller, Betsy, & Roberts, Gareth. (2018). Why some behaviors spread while others don't: A laboratory simulation of dialect contact. *Cognition* 170:298-311.

Stecker, Amelia. (2020). Investigations of the sociolinguistic monitor and perceived gender identity. *University of Pennsylvania Working Papers in Linguistics* 26(2):14.

Sumner, Meghan. (2013). A phonetic explanation of pronunciation variant effects. *The Journal of the Acoustical Society of America*. 134(1):EL26-EL32.

Sumner, Meghan., Kim, Seung Kyung, King, Ed, & McGowan, Kevin B. (2014). The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology* 4:1-13.

Tagliamonte, Sali. (2013). Comparative sociolinguistics. In J. Chambers and N. Schilling (Eds.), *The Handbook of Language Variation and Change*, 128–56.

Tamminga, Meredith. (2017). Matched guise effects can be robust to speech style. *The Journal of the Acoustical Society of America* 142(1):EL18-L23.

Vaughn, Charlotte. (2022). What carries greater social weight, a linguistic variant's local use or its typical use? Paper in the *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*.

Vaughn, Charlotte, & Kendall, Tyler. (2018). Listener sensitivity to probabilistic conditioning of sociolinguistic variables: The case of (ING). *Journal of Memory and Language* 103:58-73.

Vaughn, Charlotte, & Kendall, Tyler. (2019). Stylistically coherent variants: Cognitive representation of social meaning. *Revista de Estudos da Linguagem* 27(4):1787-830.

Villarreal, Dan, Clark, Lynn, Hay, Jennifer, & Watson, Kevin. (2021). Gender separation and the speech community: Rhoticity in early 20th century Southland New Zealand English. *Language Variation and Change*, 33:1-22.

Wade, Lacey. (2022). Experimental evidence for expectation-driven linguistic convergence. *Language* 98 (1):63-97.

Wagner, Suzanne Evans, & Hesson, Ashley. (2014). Individual sensitivity to the frequency of socially meaningful linguistic cues affects language attitudes. *Journal of Language and Social Psychology* 33(6):651-666.

Watson, Kevin, & Clark, Lynn. (2013). How salient is the NURSE~SQUARE merger? *English Language and Linguistics* 17(2):297.

Watson, Kevin, & Clark, Lynn. (2015). Exploring listeners' real-time reactions to regional accents. *Language Awareness* 24(1):38-59.

Wolfram, Walt, Childs, Becky, & Torbert, Benjamin. (2000). Tracing English dialect history through consonant cluster reduction: Comparative evidence from isolated dialects. *Journal of Southern Linguistics* 24:17-40.

Zhang, Qing. (2005). A Chinese Yuppie in Beijing: Phonological Variation and the Construction of a New Professional Identity. *Language in Society* 34:431-66.