The Journal of the Acoustical Society of America

JULY 06 2017

Matched guise effects can be robust to speech style

Meredith Tamminga

J. Acoust. Soc. Am. 142, EL18–EL23 (2017)
https://doi.org/10.1121/1.4990399

ASA
ACOUSTICAL SOCIETY OF AMERICA

Meredith Tamminga: JASA Express Letters [http://dx.doi.org/10.1121/1.4990399] Published Online 6 July 2017

Matched guise effects can be robust to speech style

Meredith Tamminga
Department of Linguistics, University of Pennsylvania, 3401-C Walnut Street, Suite 300, Philadelphia, Pennsylvania 19104, USA
tamminga@ling.upenn.edu

Abstract: When investigating how listeners evaluate the social meaning of variability in speech, researchers using the Matched Guise Technique (MGT) must decide whether to use read speech or conversational speech stimuli. An MGT experiment comparing social evaluation of /ɪŋ/ ∼ /ɪn/ variation in read and conversational speech styles found no evidence that the social evaluation of this variation differed across frame utterance styles. This suggests that use of read speech stimuli can be an appropriate methodological choice in MGT research.

[DDO]
Date Received: February 17, 2017 **Date Accepted:** June 14, 2017

1. Introduction

A commonly used method for uncovering the social meaning of variability in speech is the matched guise technique (MGT). In a typical MGT experiment, listeners hear clips of speech that (unbeknownst to them) differ minimally along some dimension. They then evaluate various social and personality characteristics of the speakers based on that speech. If listeners differ in their judgement of clips that differ in only one respect, then that evaluative difference is attributed to the linguistic difference between the clips. The MGT was pioneered by Lambert *et al.* (1960), who manipulated language choice: the same voices were heard saying the same content in both French and English. The use of English led to judgments that the speakers were significantly taller, better-looking, smarter, more dependable, etc., even though the voices and utterance content were the same.

While the MGT continues to be used to evaluate gross differences in the evaluation of varieties such as regional dialects, a further development of the paradigm is the refinement of the "guises" to differ only in a single lexical item, morpheme, phoneme, or subphonemic feature. This is most often achieved through cross-splicing the linguistic unit under investigation into frame utterances that are used across guises. The frame utterances are not simply the same words spoken by the same speaker, but repetitions of the identical audio recording. This produces stimuli that are acoustically identical except at the point of the varying item. This type of MGT experiment affords much stricter control over differences between the guises, which constitute the experimental conditions. The results from such MGT studies have been used to argue for the association of particular linguistic features with specific social meanings.

This paper reports on an experiment asking whether the differences we can evoke in social evaluation by manipulating a common English sociolinguistic variable, ING (*workin'* ∼ *working*), depend on whether the frame utterances are read speech or conversational speech. There are both methodological and theoretical issues underlying this question. Methodologically, the proliferation of MGT studies targeting small linguistic differences, including those using variant sequences (Labov *et al.,* 2011) or multiple variables (Levon and Buchstaller, 2015), motivates us to sharpen our understanding of how experimental design decisions may influence MGT results. The use of read speech versus conversational speech for stimuli is a major decision that any MGT researcher must make. It is tempting, especially for sociolinguists, to view conversational speech as the gold standard basis for understanding variation in speech. Campbell-Kibler (2009, p. 138), for example, considers the use of conversational speech for MGT "advantageous" because listener sensitivity to the unnaturalness of read speech can make social perceptions from such speech hard to generalize. However, creating read-speech stimuli in a lab is much easier than extracting appropriate stimuli from unstructured conversation, so there are real costs to the prioritization of maximally naturalistic stimuli. A researcher might reasonably wonder whether they can make meaningful progress in the study of sociolinguistic meaning without grappling with the use of conversational speech. Indeed, many well-known MGT results, such as Labov *et al.* (2011), are based on read speech.

But there are more than methodological problems at stake. Under modern sociolinguistic theories of style, variable phonetic features of speech combine in complex

ways, in real time, to constitute the linguistic construction of social meaning. The social meaning of a variable like ING, then, is neither static nor context-insensitive. In an influential line of MGT-based research on the social meaning of ING (Campbell-Kibler, 2006, 2008, 2011), Campbell-Kibler (2009, p. 141) produces extensive evidence for the "overall claim that the social contribution of (ING) is highly dependent on the other social information available in the message content, speech stream, or outside context, as well as on the overall reactions of the listeners to the speakers." Evidence for the context-sensitivity of sociolinguistic meaning is not limited to ING. For example, recent work on the perception of sexuality in both the United Kingdom (Levon, 2014) and Denmark (Pharao *et al.,* 2014) indicates that various linguistic correlates of perceived sexuality are contingent on other cues in the acoustic signal.

Furthermore, there is experimental evidence that the impact of variation on lexical access is sensitive to the embedding of variants in congruent versus incongruent forms (Gow, 2001; Sumner, 2013). Splicing a phonetic or phonological variant associated with fast-speech reduction into a word produced in clear-speech citation form can inhibit processing relative to the fully canonical form, but that inhibition seems to arise from the variant/word mismatch rather than the presence of the variant: when the variant is left in its natural whole-word form, the word is processed as rapidly as the canonical form [what Sumner *et al.* (2014) call "recognition equivalence"]. A similar effect might be expected to come into play with use of the MGT. If a non-standard variant is particularly incongruent with other acoustic/stylistic properties of the utterance it is spliced into, it may elicit stronger negative reactions than if it were presented in its "natural habitat." The experiment reported here was undertaken to test the hypothesis that evaluative differences between speech containing /ɪŋ/ and speech containing /ɪn/ will be larger in read speech than in conversational speech. This difference is expected to arise primarily from stronger disapproval (for instance, judgment of the speaker as sounding uneducated or lazy) of the /ɪn/ in read speech, where it is unusual, than in conversational speech, where it is common.

2. Methods

2.1 Participants

The experiment was administered to 49 undergraduate students from the psychology subject pool at the University of Pennsylvania and 70 United States–based workers on Prolific Academic, an online crowd-sourcing platform for academic research. Half of the Penn students and half of the Prolific Academic workers were assigned to each style condition (read speech versus conversational speech). 18 participants were excluded for reporting in the post-test demographic survey that they have not spoken English since infancy, which was adopted as a measure of native speaker status.

2.2 Procedure

The MGT experiment was presented online using the Qualtrics platform and described as a voice recognition task. Each utterance was presented on a page with an audio clip that played automatically and could be replayed by the participant. The participant was instructed to rate the voice they heard on a set of six 7-point Likert scales with the endpoints being labeled (e.g., "formal" at one side and "casual" at the other). The Likert scales targeted the following social judgments, in this order: *educated/uneducated, formal/casual, lazy/hardworking, unpretentious/pretentious, smart/stupid, unfriendly/friendly*. Of these, the first five are intended to tap specific social meanings that have previously been tied to ING (Eckert, 2008), while the latter (friendliness) is included as a control because it is not commonly proposed as a prominent social meaning of ING. The left/right orientation of each scale's endpoints was pseudorandomized (so they do not all have the same expected directionality) but held constant across all participants. Finally, the participant was presented with a binary forced choice question, "Have you heard this person's voice before during the course of the experiment?" The voice recognition question was merely a pretext intended to provide justification for the repetition of the same sentence in its two guises, thus allowing a within-subjects design for the guise effects. Participants were also given a brief demographic survey at the end of experiment and were then asked what they thought the experiment was about. Many responses to the latter mentioned accents but none indicated any awareness of the ING guise manipulation.

2.3 Stimuli

The experiment aims to compare the difference between two guises—/ɪŋ/ and /ɪn/—across two stylistic conditions: read speech and conversational speech. The conversational speech stimuli are extracted from sociolinguistic interviews in the Philadelphia

Neighborhood Corpus (PNC) (Labov and Rosenfelder, 2011), meaning that in addition to being conversational in style they also exhibit regionally marked accent features that might further license use of /ɪŋ/ (Campbell-Kibler, 2007). The extraction process worked as follows. All tokens of the variable ING occurring in a 122-person subset of the PNC had previously been identified and coded for their realization with the velar /ɪŋ/ variant or alveolar /ɪn/ variant (Tamminga, 2014). This dataset was used to identify all sets of multiple verbal ING tokens where the same speaker used the same word with the same following phoneme, and used each variant at least once within the set. For each possible ING token a short (3–8 s) audio clip of the token and surrounding utterance, containing no other instances of ING, was extracted. The sets of short clips were then subjected to a trial-and-error process of cross-splicing to produce natural-sounding pairs that differed only in the ING variant. All splices were made in Praat at a zero crossing in the waveform. Note that not all experimental stimuli were manipulated, as there were insufficient instances of comparable tokens occurring in the corpus data to find separate /ɪn/ and /ɪŋ/ variants from the same word in the same phonological context for each frame utterance.

The final set of critical stimuli contained eight unique utterances presented in two guises for a total of 16 clips to be judged by listeners. An additional 16 clips containing no instances of ING, two each from eight talkers who did not contribute critical stimuli, were included as fillers. In four cases the two utterances from a filler talker were identical, and in the remaining four the same talker contributed two distinct utterances. The critical and filler utterances were pseudorandomized within two blocks so that no critical utterance pairs occurred fewer than 11 trials apart. If an utterance appeared in its /ɪŋ/ guise in block one, it occurred in its /ɪn/ guise in block two, and vice versa. The ING guises were counterbalanced across blocks with talker gender and whether or not the utterance was manipulated. The block design was not apparent to the participants.

The frame-style manipulation was done with a between-subjects design, so that participants were randomly assigned to either the read-speech or conversational-speech condition. The read-speech condition was created by recruiting a new set of model talkers (mostly linguistics graduate students at the University of Pennsylvania) to read transcriptions of the materials from the conversational speech stimuli. Female readers were matched to the utterances from female original speakers, and likewise for male readers/speakers. Readers were allowed to listen to the original conversational speech versions of the utterances they were reading, and were explicitly asked to follow the basic intonational contours of the original (for example, using question intonation if the original speaker did so) while not imitating the accent or trying to sound spontaneous. Readers provided three or four instances of each utterance, and only the instance that was subjectively judged to best meet these goals was selected for use in the experiment. Readers providing critical stimuli read the utterances in both guises, so that cross-splicing could be done parallel to the original conversational-speech stimuli. Spectrograms for the word "walking" in both its original alveolar form and spliced velar form in the read speech condition are shown in Fig. 1. Note the velar pinch in the velar spectrogram.

3. Results

Data analysis was done using mixed effects linear regression with the lme4 package (Bates *et al.,* 2015) in R (R Core Team, 2015). For each critical pair of utterances for each participant, a guise difference score was calculated: the participant's rating of the utterance in the /ɪn/ guise minus the participant's rating of the utterance in the /ɪŋ/ guise. This guise difference score was then adopted as the dependent variable in a Generalized Linear Mixed Model (GLMM) to ask whether the magnitude of the guise difference was affected by the speech style. The fixed effect predictors in the regression model were participant gender, model talker gender, the interaction of participant and talker gender, whether the participant was recruited at the University of Pennsylvania or on Prolific Academic, which guise in the pair was presented first, type of speech

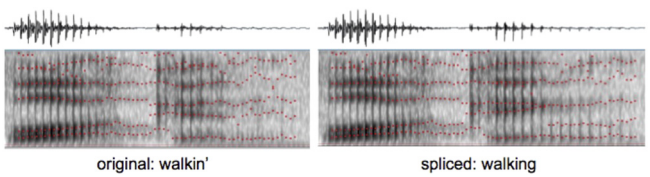

Fig. 1. (Color online) Waveforms and spectrograms for the word "walking" in two guises in read speech.

(read or conversational), which social scale the rating was to (e.g., formal/casual), and the interaction of speech type and social scale. All categorical predictors were sum-coded except the manipulation of speech type, which was treatment-coded with conversational speech as the reference level, and social scale, which was treatment-coded with the unfriendly–friendly scale as the reference level. Additionally, random intercepts were included for participant identity and utterance identity. The model fit is presented in Table 1. The *t* value is the ratio of the coefficient to the standard error, and as a rule of thumb reflects a statistically significant effect when it is over 2.

Of the control predictors (gender, participant source, presentation order), there is a significant effect only of participant location: students at Penn have larger guise differences than workers on Prolific Academic. The near-zero intercept in the model indicates that for conversational speech (averaging across participants and orders) there is no significant difference between /ɪŋ/ and /ɪn/ ratings on the friendliness scale, as expected. There is also not a significant main effect of read speech. Recall that the dependent variable here is guise difference, so this does not indicate that participants rate read and conversational speech the same (in fact, the read speech condition has overall higher ratings). When we turn to the main effects of the various social scales, we see that education, formality, and pretentiousness all differ significantly from friendliness in the guise differences they evince. The /ɪn/ guise is rated as more uneducated, more casual, and less pretentious. These effects can be seen visually in the left-hand facet of Fig. 2, which presents simple descriptive statistics for the data underlying the multivariate model. The sensitivity of guise difference to social scale suggests that the MGT manipulation of ING was successful in shifting social judgments of the model talkers. Turning to the interaction of speech style and social scale, however, there is no evidence to support the hypothesis of larger guise differences in read speech. This lack of interaction is also visible in the generally similar pattern of guise differences across the left and right facets of Fig. 2.

4. Discussion

The MGT experiment reported here aimed to test the hypothesis that the social evaluation of ING variability would be larger when the variation was presented in read speech than when it was presented in conversational speech. The reasoning behind this hypothesis was that while both the /ɪn/ and /ɪŋ/ variants are compatible with conversational speech, use of the /ɪn/ variant is particularly incongruent with read speech and should therefore elicit stronger negative judgements in that stylistic context. The results of the experiment do not support the hypothesis. While the guise manipulation between /ɪn/ and /ɪŋ/ successfully shifted participants' judgments of the model talkers on the expected social scales, these shifts took place to similar degrees regardless of whether the frame utterances were read speech or conversational speech: none of the interaction terms between social scale and speech style were statistically significant.

Table 1. GLMM predicting /ɪn/–/ɪŋ/ guise difference across social scales in read speech and conversational utterances, N = 4848.

	Estimate	SE[a]	*t* value
(Intercept)	−0.02	0.07	−0.25
Female participant	−0.01	0.02	−0.49
Female talker	0.02	0.02	0.76
Penn participant	0.07	0.03	2.12
/ɪn/ guise first	−0.004	0.04	−0.09
Read speech	0.05	0.09	0.56
Educated–uneducated	0.18	0.09	2.08
Formal–casual	0.23	0.09	2.56
Lazy–hardworking	−0.03	0.09	−0.32
Smart–stupid	0.12	0.09	1.38
Unpretentious–pretentious	−0.19	0.09	−2.15
Female participant: Female talker	−0.02	0.02	−1.33
Read speech: Educated–uneducated	−0.15	0.12	−1.25
Read speech: Formal–casual	−0.01	0.12	−0.11
Read speech: Lazy–hardworking	−0.11	0.12	−0.91
Read speech: Smart–stupid	−0.04	0.12	−0.37
Read speech: Unpretentious–pretentious	−0.00	0.12	−0.004

[a]Standard error (SE).

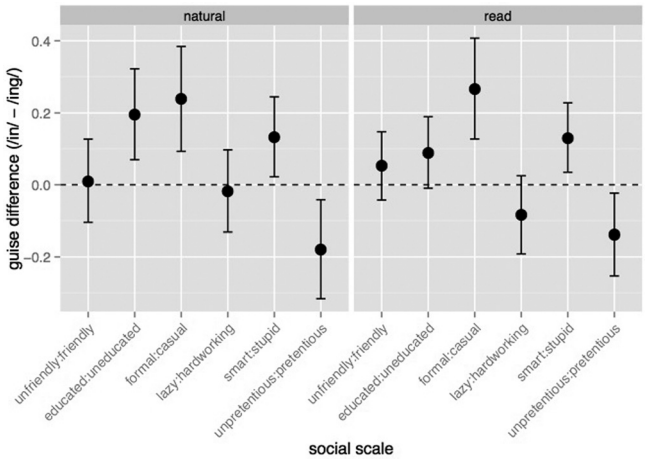

Fig. 2. Means and 95% confidence intervals for guise differences by speech style and social scale.

Methodologically, this null result is encouraging for researchers using MGT experiments. It suggests that the use of easy-to-create read speech stimuli is a reasonable methodological choice, one which can successfully elicit social evaluation differences comparable in direction and magnitude to those from naturalistic stimuli. Two cautions should be offered with this point. The first is that the read speech stimuli used here were modeled on naturally occurring utterances in their lexical and grammatical choices; it is possible that use of more structurally formal written English sentences would have given rise to larger guise effect differences. The second is that other sociolinguistic variables besides ING might be more sensitive to the difference between read speech and conversational speech. It would be misguided to respond to the results of this experiment by withdrawing attention to stylistic considerations in experimental design for social speech perception research. With sufficient caution, however, use of read speech stimuli can be an appropriate methodological choice.

The largely parallel ING guise effects in read and conversational speech that were observed here are not incompatible with previous research demonstrating the context-sensitivity of social meaning. They do, however, suggest that such sensitivity is not infinite; social meanings can be robust to some contextual differences. In this study, it appears that the core social meanings of ING are quite stable across utterance pairs that have the same content but different phonetic styles and voice characteristics. Such stability points to content-based differences—*what* people are saying, not just how they are saying it—as a potentially fruitful area for further inquiry on the dynamism of social meaning.

Acknowledgments

Thanks to Elisha Cooper for her assistance with data collection.

References and links

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**(1), 1–48.

Campbell-Kibler, K. (**2006**). "Listener perceptions of sociolinguistic variables: The case of (ING)," dissertation, Stanford University, Stanford, CA.

Campbell-Kibler, K. (**2007**). "Accent, (ING), and the social logic of listener perceptions," Am. Speech **82**(1), 32–64.

Campbell-Kibler, K. (**2008**). "I'll be the judge of that: Diversity in social perceptions of (ING)," Lang. Soc. **37**(05), 637–659.

Campbell-Kibler, K. (**2009**). "The nature of sociolinguistic perception," Lang. Var. Change **21**, 135–156.

Campbell-Kibler, K. (**2011**). "The sociolinguistic variant as a carrier of social meaning," Lang. Var. Change **22**, 423–441.

Eckert, P. (**2008**). "Variation and the indexical field," J. Socioling. **12**(4), 453–476.

Gow, D. W., Jr. (**2001**). "Assimilation and anticipation in continuous spoken word recognition," J. Mem. Lang. **45**, 133–159.

Labov, W., Ash, S., Ravindranath, M., Weldon, T., Baranowski, M., and Nagy, N. (**2011**). "Properties of the sociolinguistic monitor," J. Socioling. **15**(4), 431–463.

Labov, W., and Rosenfelder, I. (**2011**). "The Philadelphia Neighborhood Corpus of LING 560 studies, 1972-2010," NSF contract 921643.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., and Fillenbaum, S. (**1960**). "Evaluational reactions to spoken languages," J. Abnorm. Soc. Psychol. **60**(1), 44–51.

Levon, E. (**2014**). "Categories, stereotypes, and the linguistic perception of sexuality," Lang. Soc. **43**, 539–566.

Levon, E., and Buchstaller, I. (**2015**). "Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing," Lang. Var. Change **27**, 319–348.

Pharao, N., Maegaard, M., Moller, J. S., and Kristiansen, T. (**2014**). "Indexical meanings of [s+] among Copenhagen youth: Social perception of a phonetic variant in different prosodic contexts," Lang. Soc. **43**, 1–31.

R Core Team (**2015**). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).

Sumner, M. (**2013**). "A phonetic explanation of pronunciation variant effects," J. Acoust. Soc. Am. **134**(1), EL26–EL32.

Sumner, M., Kim, S. K., King, E., and McGowan, K. B. (**2014**). "The socially weighted encoding of spoken words: A dual-route approach to speech perception," Front. Psychol. **4**, 1015.

Tamminga, M. (**2014**). "Persistence in the production of linguistic variation," dissertation, University of Pennsylvania, Philadelphia, PA.