# Perceptual similarity across multiple sociolinguistic variables

Elizabeth D. Casserly[1,2]

[1]Dept. of Linguistics, Indiana University, 322 Memorial Hall, 1021 E. 3rd St., Bloomington, IN 47405

[2]Dept. of Psychological & Brain Sciences, Speech Research Laboratory, Indiana University

**Abstract**

Although variationist studies of sociolinguistic performance have greatly enhanced our understanding of speakers' use of linguistic variables and their correspondence with social and contextual factors, these works have largely overlooked the perceptual complement of the production behavior they capture. This paper reports the results of a study which uses multi-dimensional scaling (MDS) to model listeners' similarity judgments of tokens of the word *going* produced by male and female talkers of Southern and non-Southern dialects of American English, with and without g-dropping and vowel quality differences in the suffix *–ing*. Listeners were found to weigh heavily the more fluid factors (g-dropping and vowel quality) and nearly ignore the more static characteristics (talker dialect and gender).

## 1. Introduction

Empirical studies of variation in speakers' use of language form one of the cornerstones of sociolinguistic inquiry. Essentially begun and defined by William Labov, the variationist approach uses data collected from sociolinguistic interviews or public speech corpora such as radio programs as evidence of how real speakers use language in actual situations (e.g. Labov, 1966; 1972). Researchers analyze the data to get a sense of the social patterns underlying the patterns of linguistic use. Outside of the sociolinguistic tradition, variation between alternate forms in language has been described as "free," or unrestricted by traditional syntactic, phonological or semantic factors (Chomsky and Miller, 1963). As variationist sociolinguistics has demonstrated, however, speakers' choice of alternating or competing forms is far from unrestricted; social factors such as a speaker's age, gender, socio-economic status (SES) and dialect or their perception of a situation's formality or tone exert strong influence on language use (Labov, 1972). By documenting the speech of individuals spanning the width and breadth of these social variables, variationist studies have yielded a great deal of insight into the regular, non-random patterns underlying the wide variety of observed language use (see Wardhaugh, 2006, p.146-190 for an overview).

While our understanding of the tendencies governing speakers' production of specific linguistic variants has greatly improved as a result of the variationist approach, little is currently known about how listeners perceive and use sociolinguistic information in their social and linguistic interactions. The majority of the literature focuses on studies that allow researchers to make claims about the correlation between linguistic production variables such as word choice (e.g. use of *ain't*) or pronunciation (e.g. [r]-full versus [r]-less) and speaker characteristics such as SES or social-group affiliation. Attempts to go beyond correlation into claims about what particular linguistic variants signal to *listeners* about social factors also go beyond what the production data alone can support. Typical data from speakers' productions provides no evidence for listener behavior, either that listeners attend to the identified production variables or that they use the variables to draw any straightforward conclusions about speakers' social status. Since speech is inherently a two-way street, involving the behavior of both a speaker and a listener for every act, our understanding of sociolinguistic variation cannot be complete until its role in perception is addressed.

Some sociolinguists are beginning to undertake the research necessary to fill this gap. Campbell-Kibler, for example, recently completed a series of studies investigating the kinds of judgments listeners draw about speakers after hearing short samples of their speech (2006, 2007, 2009). She took excerpts of speech from sociolinguistic interviews with several talkers from North Carolina and several others from California, and spliced the tokens so that they could appear with either full or g-dropped pronunciations of *ing* words (i.e. *jumping* could appear as either [dʒʌmpɪŋ] "jumping" or [dʒʌmpɪn] "jumpin'"). She played one version or the other to groups of naïve listeners and asked them general questions about their impressions of the speakers' personalities and regional backgrounds. Although she could have manipulated many different linguistic features in her stimuli, Campbell-Kibler chose to investigate listeners' interpretation of g-dropping for two reasons. First, its pattern of use has been extremely well documented in the variationist literature (i.e. starting with Fischer, 1958); and second, it is a linguistic feature which crosscuts almost every social group, appearing to some degree in the speech of nearly every speaker of American English (and other varieties) (Campbell-Kibler, 2007). She reasoned, therefore, that her listeners would be neither closely clued into any particular variety of English by the g-dropping variable nor unfamiliar with its occurrence in the language. And as expected, she found that the g-drop variable significantly affected listeners' perceptions of the speakers, controlling for all other factors (Campbell-Kibler, 2006, 2007, 2009).

Although the studies of Campbell-Kibler and others (e.g. Eckert, 2008) constitute a promising start to the study of the perception of sociolinguistic variables, much work remains to be done. Below, I describe one early attempt to further this line of research using data analysis methods common to psychological research on human categorization and perceived similarity of categories or individual items (Shepard, 1962; Tversky & Gati, 1978; Kruskal & Wish, 1978).

## 2. Perceptual Experiment

Previous research investigating listeners' use of sociolinguistic variables in perception has asked listeners to make explicit judgments about people, such as where they are from or what their personalities are like, based on brief samples of their speech (e.g. Campbell-Kibler, 2007). In such methodology, however, it is possible for the questions asked to influence listeners' responses. Without intentionally doing so, researchers may lead listeners to evaluate speakers in particular ways, or pay attention to particular variables. Although care has been taken to avoid these problems in previous work, the need remains for a maximally objective and general examination of listeners' perceptual behavior. Instead of making overt judgments, therefore, listeners in this experiment were asked to rate the similarity of two instances of short (word-length) tokens of recorded speech. No explicit criteria for similarity were given or implied so that the listeners' task would be maximally general and open-ended, without any possible leading from the investigator.

Since so little is known about the sociolinguistic variables listeners attend to in perception, I wanted to offer a variety of possible factors which they could use as their basis for judging the similarity of the speech they heard. The semantic content of the speech was held constant; listeners heard different tokens of a single word, *going*, varying on four "extra-semantic" factors. For largely the same reasons as it was included in the work of Campbell-Kibler above, I chose to manipulate the g-dropping present in *going*. Since the generality of these results are of such strong interest, however, I also included a less common variant of *ing* word realization, which includes a high tense vowel [i] before the final nasal (see de Wolf, 1988; Woods, 1979). In addition to these g-drop and vowel quality factors, I also controlled the gender and dialect of the recorded speakers, such that tokens from male and female speakers who spoke either general American English or a Southern dialect were included. Given speech that fully crossed the four factors of talker gender, talker dialect, g-drop and vowel quality, how would listeners judge similarity? Which factors would be most salient or vital, and would the choice vary across individual listeners?

### 2.1 Subjects

Speech was recorded from four native speakers of American English, one male and one female speaker of a Southern dialect (Southern Appalachia) and one male, one female speaker of non-Southern dialects (the female was from rural Pennsylvania and the male from Massachusetts – neither exhibited strong phonetic features associated with specific regions such as Philadelphia or Boston). All speakers were between 21 and 26 years of age and current residents of Bloomington, Indiana.

Five native speakers of American English were recruited as participants in the subsequent perceptual experiment, none of whom also acted as speakers during the creation of the stimuli. These listeners fell into the same age range as the talkers and were also current residents of Bloomington. Three listeners were raised in the Midwest and two in the South (North Carolina and Oklahoma). No substantial differences across listeners, due to dialect or other factors, were observed. All participants, listeners and talkers, were volunteers receiving no compensation for their participation.

### 2.2 Materials & Stimulus Collection

In order to generate the stimuli used in the perceptual experiment, the four talkers were recorded via head-mounted microphone reading a word list from a computer screen in a sound-attenuating booth. They produced each word three times, with desired pronunciation noted orthographically for g-dropping and vowel quality. These variants were discussed with speakers before the recording session and practiced with the investigator so that they accurately reflected the following IPA transcriptions: [goɪŋ] (lax vowel, no g-drop), [goɪn] (lax vowel, dropped-g), [goiŋ] (tense vowel, no g-drop) and [goin] (tense vowel, dropped-g). Producing precisely these variants of *going* neutralized some of the natural variation present in the speakers' productions (due to gender, dialect, SES, etc.), but variations in vowel quality, speaking rate and other acoustic phonetic factors remained. These factors were each potential dimensions, in addition to g-dropping and vowel quality, that could serve as the basis for listeners' similarity judgments. Certain acoustic, phonetic and sociolinguistic aspects can be expected to co-vary according to the speakers' gender (e.g. fundamental frequency or pitch) or dialect (e.g. vowel formant spacing); those factors will be loosely grouped as "dialect" and "gender" in the analysis below.

**Table 1.** Stimulus values for each of four sociolinguistic variables. M = male, F = female; tn = tense, lx = lax; N = non-Southern, S = Southern; g = non-g-dropped, n = g-dropped

| Stimuli | Talker Gender | Vowel Quality | Talker Dialect | G-Dropping |
|---|---|---|---|---|
| 1 | M | tn | N | g |
| 2 | M | tn | N | n |
| 3 | M | lx | N | g |
| 4 | M | lx | N | n |
| 5 | F | tn | N | g |
| 6 | F | tn | N | n |
| 7 | F | lx | N | g |
| 8 | F | lx | N | n |
| 9 | M | tn | S | g |
| 10 | M | tn | S | n |
| 11 | M | lx | S | g |
| 12 | M | lx | S | n |
| 13 | F | tn | S | g |
| 14 | F | tn | S | n |
| 15 | F | lx | S | g |
| 16 | F | lx | S | n |

The most natural-sounding token of the three repetitions of each variant, as determined by the investigator, was then digitally edited (removed from surrounding noise) for use as a perceptual stimulus. Once the four tokens, one for each pronunciation variant of *going*, were selected for each subject, they were digitally leveled so that the mean amplitude of the sound was consistent across speakers and tokens. Leveling ensured that listeners could not use volume as a means of judging similarity. The characteristics of the 16 stimuli are summarized in Table 1.

### 2.3 Methods: Perception Task

Participants in the perception experiment were seated in front of Macintosh computers and fitted with headphones. The task was explained verbally by the investigator, with written explanation of the instructions given again on the computer screen before the task began. For each trial in the experiment, listeners heard a pair of stimuli chosen at random from a list of all possible pairs. One second intervened between the two audio files. After hearing the second sound file, listeners pressed a numeric key from 1 to 9, indicating their rating of the similarity of the two tokens in the trial. "1" indicated that the tokens were "not at all similar," while "9" meant that they perceived the two tokens as "nearly identical." No trial consisted of two repetitions of an identical stimulus, and the participants were aware of this restriction so that they would feel comfortable using a “9” rating even if samples were not identical.

Over the course of the experiment, each stimulus was paired with every other stimulus (not including itself) five times, generating five separate judgments of the similarity of every stimulus to every other from each listener. The order of presentation of the tokens in a pair was fixed such that one order (e.g. 1-2) was heard twice and the other (2-1) was heard three times. Every subject heard the same pairs, with the same number of each potential ordering within pairs, in a random order over the course of the experiment. Stimulus presentation and randomization, as well as the recording of participants’ responses, was done with PsyScript (Slavin, 2007). Stimuli were presented at 73 dB, and there were a total of 600 trials. The experiment lasted approximately 45 minutes.

### 2.4 Data Analysis: Multidimensional Scaling

The data from the perceptual similarity judgments were analyzed using multi-dimensional scaling (MDS). This methodology creates multi-dimensional maps of experimental stimuli based on translation of similarity to distance such that the higher the similarity rating between two stimuli, the smaller the distance between them in the map and vice versa. MDS is a staple of the psychological literature on categorization and perceptual grouping (Shepard, 1962; Tversky & Gati, 1978; Kruskal & Wish, 1978). Dimensions on the maps correspond to factors or scales which were meaningful to the participants who responded to the experimental stimuli. Given a similarity-rating task using animal names, for example, a map with two dimensions has been found to capture most people’s similarity judgments extremely well, with one dimension placing stimuli (the various animals) one a scale reflecting their status as a predator or prey animal and the other reflecting the animals’ size (e.g. “giraffe” is positioned on the map such that it is both a prey animal and a large animal, “mouse” is prey but small, “tiger” is a large predator and “cat” is a small predator) (Romney et al., 1993). Once a stimulus map is derived, the interpretation of its dimensions constitutes the bulk of the analytical challenge in MDS.

The goodness of fit of an MDS map, or its ability to capture the obtained similarity judgments, is assessed by a measure of “stress.” Any map solution will deviate from the raw data to some degree, increasing its stress, but lower stress values correspond to better fits. Because maps are generated to capture relative similarities among a group of stimuli, the precise values associated with its dimensional scales are irrelevant. MDS algorithms assign scales to dimensions for comparative purposes, but maps are general to any scaling that preserves relative distance between points.

In order to make use of MDS analysis, a matrix was created of the average similarity rating for each possible stimulus pair across all repetitions and all five listeners. A small portion of this matrix is given in Table 2. The matrix was then used to derive a multi-dimensional stimulus map via the ALSCAL scaling algorithm (Shepard, 1962).

**Table 2.** A portion of the similarity rating matrix generated by the five listeners' ratings from 1—9, expressed as a percentage (1 = 100% similar, average rating of 9.0)

| Stimulus # | 1 | 2 | 3 | 4 | 5 | 6 | … |
|---|---|---|---|---|---|---|---|
| **1** | 1 | | | | | | |
| **2** | 0.728889 | 1 | | | | | |
| **3** | 0.431111 | 0.444444 | 1 | | | | |
| **4** | 0.421296 | 0.373333 | 0.68 | 1 | | | |
| **5** | 0.737778 | 0.582222 | 0.511111 | 0.502222 | 1 | | |
| **6** | 0.62963 | 0.56 | 0.453333 | 0.44 | 0.813333 | 1 | |
| **…** | | | | | | | |

## 3. Results: MDS Solution & Interpretation

The optimal scaling solution consists of a 3-dimensional map with stress = 0.108 (a fit within the conventions of significance). The solution also accounts for a high percentage of the variance in the data (r-squared = 0.905). Each of the 3 dimensions present in the solution had a clear and unambiguous interpretation, reflecting one of the sociolinguistic variables manipulated in the experiment. This reflection is such that the first dimension splits the tokens according to their vowel quality, the second separates tokens with and without g-dropping, and the third dimension corresponds to talker dialect, splitting the Southern speakers from the speakers of the relatively unmarked dialect. Talker gender was not found to be a salient dimension under any MDS solution, which is reflected by the fact that the optimal stimulus map contained only 3 dimensions. Perceptual space maps that "flatten" one dimension at a time for more convenient visualization of the solution are given in Figures 1—3.

Figure 1 shows dimension 1 (vowel quality) on the x-axis and dimension 2 (g-dropping) on the y-axis; dimension 3 (dialect) is collapsed. All of the stimuli containing lax vowels (signaled by "lx" in their label) are on the left, and all of the tense vowel stimuli ("tn") are on the right. Simultaneously, the majority of the non-g-dropped stimuli (marked with "g") are on the bottom of the graph, while their g-dropped counterparts tend to be on the top. In Figure 1, tokens produced with lax vowels are marked by diamonds, those with tense vowel tokens are marked by circles and g-dropped tokens with either vowel quality are indicated by stripes within the markers. Exceptions to the top/bottom split are ringed – there are no exceptions to the left/right vowel quality split.

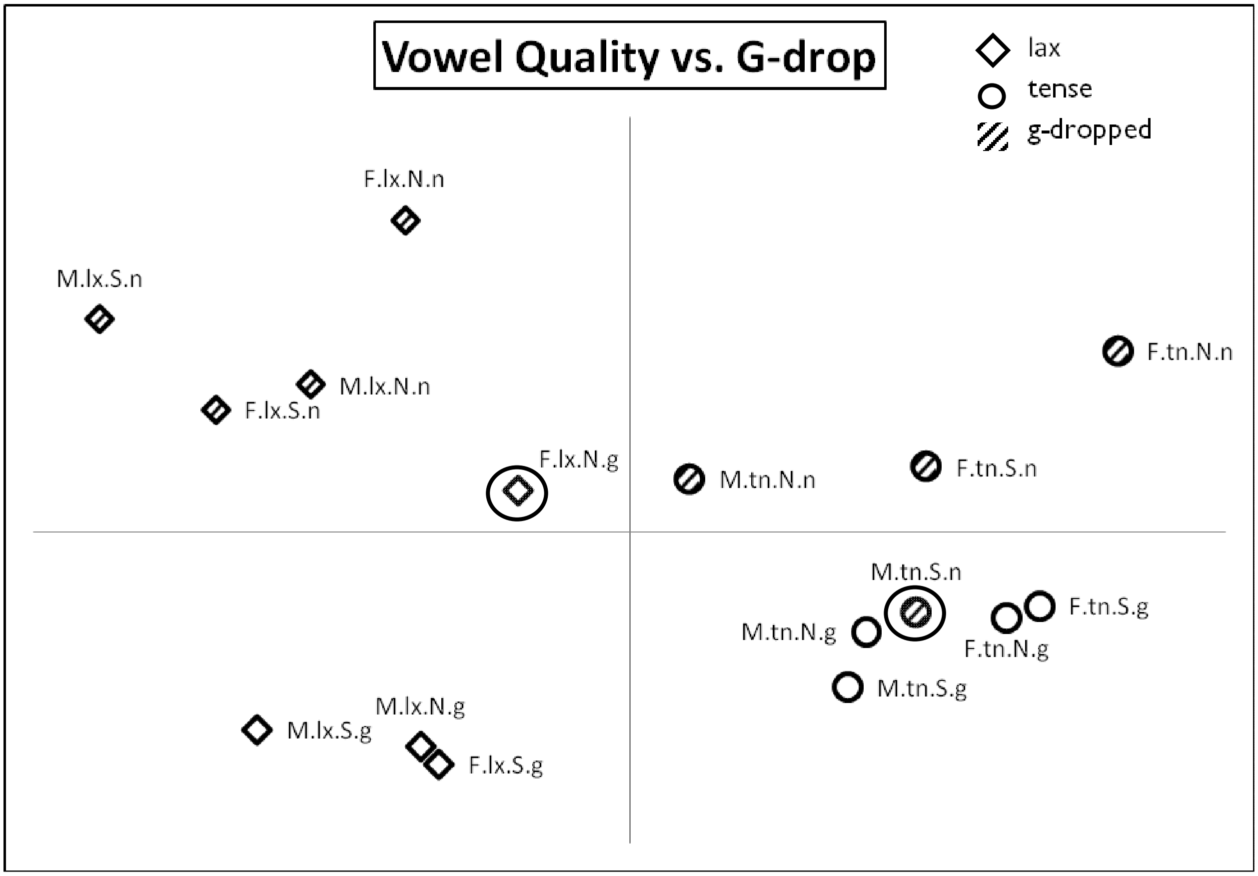


**Figure 1.** Dimensions 1 and 2 of the 3-D MDS solution. Vowel quality runs left to right on the x-axis such that all tense vowels (marked "tn") are on the right and lax vowels ("lx") are on the left. The g-dropping dimension runs top to bottom, with g-dropped stimuli ("n") on top and non-g-dropped stimuli ("g") on bottom. Exceptions to the top/bottom split are marked with rings. Tense stimuli are circles, stimuli with lax vowels are diamonds, and g-dropping is represented by stripes.

Figure 2 shows dimension 1 (vowel quality) plotted against dimension 3 (talker dialect) on the y-axis; dimension 2 (g-dropping) is collapsed. Just as in Fig. 1 above, all the lax vowel stimuli are on the left while the tense vowel stimuli are on the right, corresponding to the vowel quality dimension. In this Figure, however, the majority of the non-Southern ("N") stimuli are towards the top of the graph while the stimuli from the Southern talkers ("S") are towards the bottom. The four exceptions are marked with rings. Once again, the shape of each marker reflects its vowel quality; those that are also Southern are filled with stripes.

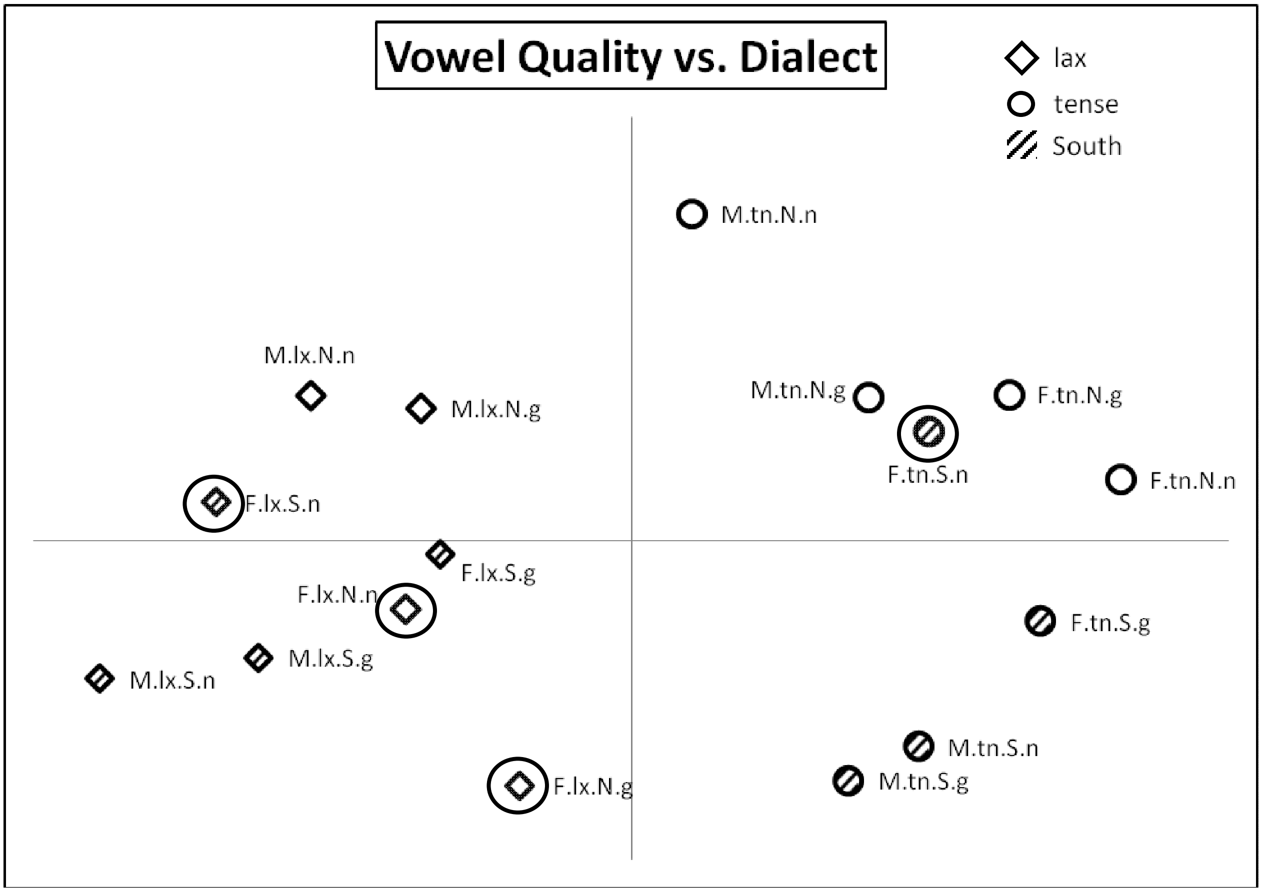


**Figure 2.** Dimensions 1 and 3 of the 3-D MDS solution. Vowel quality runs left to right on the x-axis such that all tense vowels (marked "tn") are on the right and lax vowels ("lx") are on the left. The talker dialect dimension runs top to bottom, with stimuli from non-Southern talkers ("N") on top and those from Southern speakers ("S") on bottom. Exceptions to the top/bottom split are marked with rings. Tense stimuli are circles, stimuli with lax vowels are diamonds, and stimuli from Southern talkers are represented by stripes.

Finally, Figure 3 shows dimension 3 (dialect) on the x-axis and dimension 2 (g-dropping) on the y-axis. This Figure is the most difficult to interpret because both dimensions have exceptions. Particularly since they have both previously been identified, however, it is possible to see that the stimuli from Southern talkers tend to be on the left and the non-Southern talkers on the right (barring ringed exceptions), while the g-dropped stimuli tend to be on top and non-g-dropped stimuli mostly appear on the bottom half of the graph. To help with interpretation, a diamond (Southern) and circle (non-Southern) marker are positioned on the appropriate ends of the x-axis. Although not as crisp as those views in Figures 1 and 2, this final view of the 3-dimensional solution confirms the interpretation of the weaker dimensions as reflecting g-dropping and talker dialect, respectively.

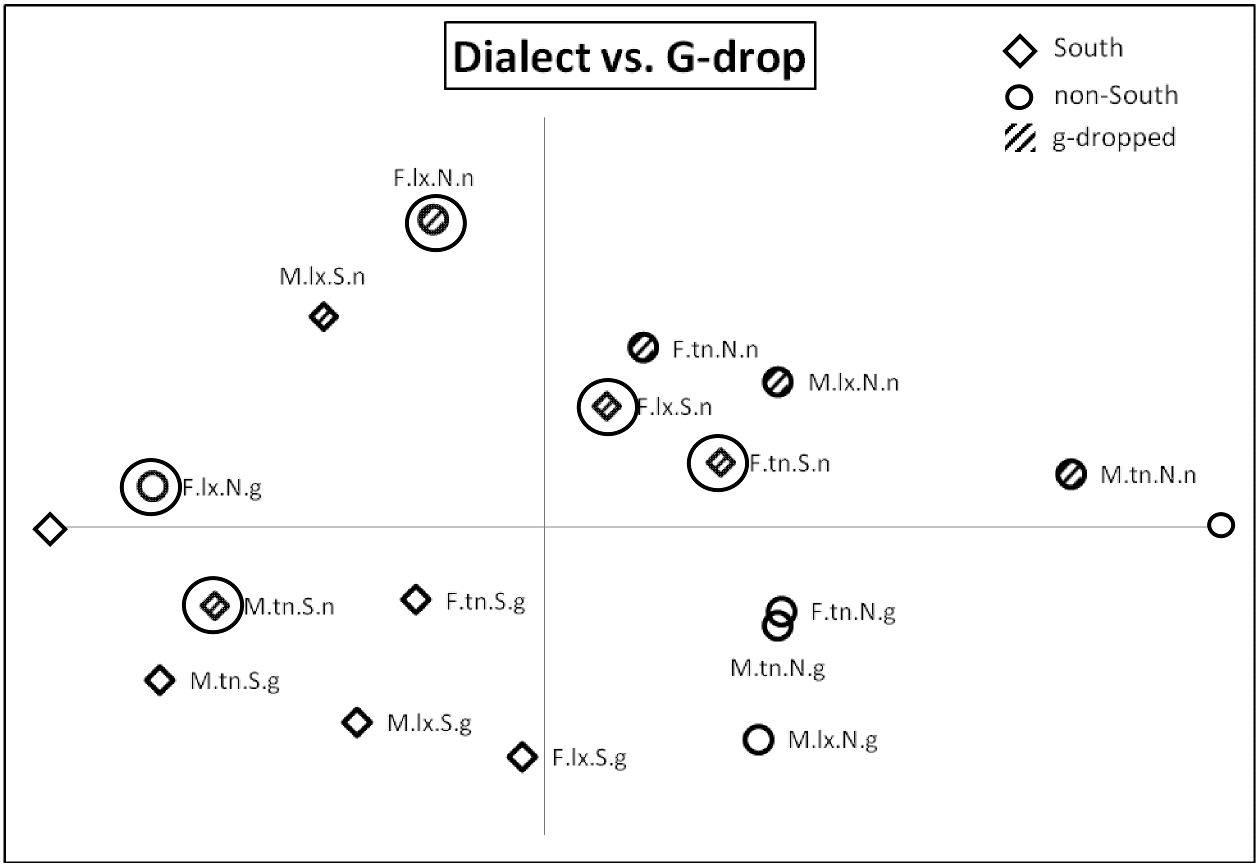


**Figure 3.** Dimensions 2 and 3 of the 3-D MDS solution. Talker dialect runs left to right on the x-axis such that stimuli from non-Southern talkers ("N") tend to be on the right and those from Southern speakers ("S") tend to be on the left. G-dropping runs from top to bottom as in Fig. 1, such that g-dropped tokens ("n") are on top and non-dropped tokens ("g") are on bottom. Exceptions to the splits are marked with rings. Stimuli from Northern talkers are circles, those from Southern talkers are diamonds, and g-dropping is indicated by stripes.

Although the interpretation of MDS solution dimensions constitutes a crucial part of their analysis, the optimal dimensionality of a solution is ultimately a subjective decision, as increasing the number of dimensions always results in a model with a better fit. The benefit of added dimensions must therefore be taken in the context of the added theoretical cost they incur: additional dimensions that do not have a clear interpretation or that increase fit only marginally probably do not capture a practically significant aspect of the similarity structure.

For these data, although there were four sociolinguistic variables manipulated in the stimuli, a three-dimensional solution was selected. The value of stress obtained for solutions of increasing dimensionality (represented in the stress plot in Figure 4) decreased gradually, without a strong "elbow," or sharp leveling off of the benefit gained by adding dimensions to the solution. The interpretability of dimensions, however, played a decisive role in the selection of a 3-dimensional optimal solution. In the 3-dimensional case, the axes of the similarity map clearly corresponded to sociolinguistic variables manipulated in the experiment, splitting contrasting tokens about the origin. In the 4-dimensional solution, two axes were without clear corollaries in the stimuli; they seemed to be splitting up the work done by one axis in the 3-dimensional solution. In light of this, the 3-dimensional solution seemed to capture the similarity structure perceived by listeners most accurately.

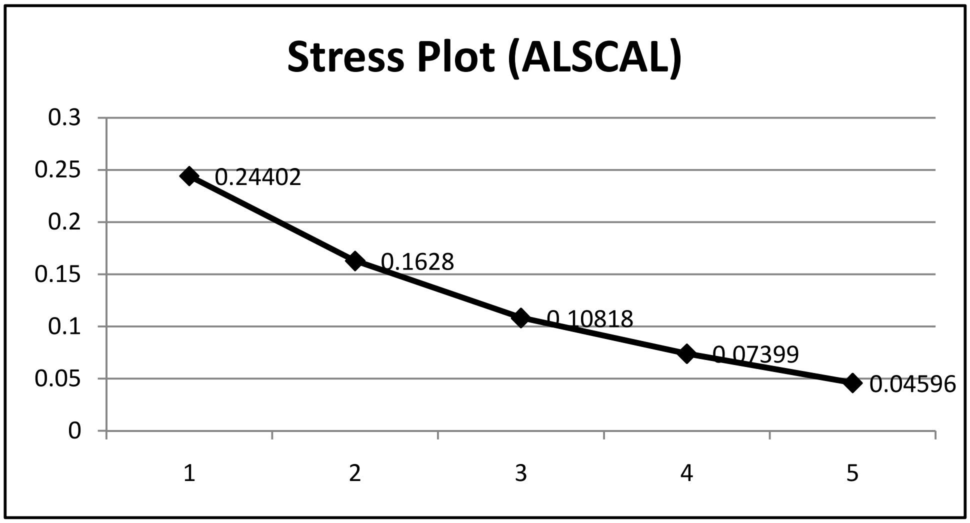


**Figure 4**. Stress plot of the stress values (measures of fit) obtained for MDS solutions with 1—5 dimensions. No clear drop in fit benefit with any one dimensional addition is found.

Given the 3-dimensional solution, the ordering of the dimensions from first to third reflects their relative strength or perceptual saliency for the listeners' similarity judgments. Across subjects, vowel quality was the most salient factor used in rating similarity, followed by g-dropping, then talker dialect. More salient factors also have the fewest stimuli which constitute exceptions to general map placement: as Figures 1 through 3 show, there are no exceptional stimuli on the vowel quality dimension (i.e. no point is on the "wrong" side of the origin on that dimension, given its interpretation), two exceptions to the g-dropping dimension and four exceptions (of 16 stimuli) on the dialect dimension.

## 4. Discussion

In this experiment, naïve listeners were asked to rate similarity on a scale from 1 to 9 for pairs of tokens of the word *going* that crossed the values of four different sociolinguistic variables: talker gender, talker dialect, the vowel quality of the suffix (tense [i] or lax [ɪ]) and the dropping of the final "g" ("going" versus "goin'"). The average similarity ratings were then used in a multi-dimensional scaling (MDS) analysis to derive the dimensions or factors which the listeners used in making their similarity judgments, and with what relative strength.

The derived 3-dimensional scaling solution provided a good fit to the data and accounted for just over 90% of the variance observed in listeners' similarity judgments. The fact that a *three*-dimensional solution proved optimal for a stimulus set which varied along *four* factors signifies that listeners essentially ignored one of the factors when they decided how strongly to rate the similarity of pair of stimuli. Perhaps surprisingly, talker gender appears to be the factor that listeners in this experiment ignored. Even though interpretation of scaling dimensions and choosing the number of dimensions is in some sense subjective, in no solution were stimuli split according to talker gender along a scaling dimension. The absence of gender information from the perceptual similarity structure therefore cannot be a property of the specifics of a 3-dimensional solution.

The lack of gender effect goes against a substantial body of literature demonstrating that listeners are sensitive to gender differences in speech. Listeners can make very accurate gender judgments on the basis of speech recordings (Lass et al., 1976) and perceived speaker gender influences even low-level processes like phonetic categorization (Munson et al., 2006). Many of these and related studies, however, either overtly draw attention to gender or demonstrate its covert influence on other cognitive processes. In this experiment, listeners made overt judgments, but about similarity – a fluid and open-ended construct. The results show that, in this one case of conscious open-ended comparison, our listeners chose to ignore speaker gender. They do not suggest that the listeners could not perceive gender or that their perception was not in any way influenced by gender; just that the overt judgments were not differentiated by gender.

Of the dimensions which did have a significant impact on similarity judgments, the vowel quality manipulation proved most salient to the listeners, followed by whether or not the token was g-dropped and finally by the dialect of the talker. Prior to conducting the study, there was some question as to whether regional dialect could even be reliably conveyed in a speaker's production of a single highly-frequent word such as *going*. As this dimension does appear to effectively separate tokens in the MDS solution, however, dialect information must be present even in these small samples of speech. Perhaps with more extreme realizations of dialect the perceptual salience would be stronger relative to the vowel quality and g-dropping features.

In any case, it is striking that those factors over which speakers have the most control, or which change most readily with stylistic and formality variations in production studies, are those which listeners weighed most heavily in their perceptual similarity judgments. No specific instructions for basing similarity were given; these features were attended to spontaneously by our listeners. Impressionistically, talker gender is a highly salient feature of these tokens and speech in general: it is not the case that there were insufficient acoustical differences between tokens from male and female talkers for listeners to reliably detect gender. Rather, participants seem to have actively disregarded these differences in their rating of similarity.

Similarly, talker dialect is a sociolinguistic marker that has been receiving increasing amounts of attention in perceptual research (e.g. Clopper, 2004a, 2004b; Johnson, 2005). Just as with gender, we know that dialect markers or regional accents affect speech perception (Niedzielski, 1999) and listeners are able to classify and group both familiar and unfamiliar dialects, overtly and tacitly making use of their accentual features (Clopper, 2007; Tamati, in press). Yet talker dialect played the least active role in our listeners' judgments of similarity, other than the absent effects of gender. These two factors have in common their entrenchment in speakers' language – for the most part, speakers cannot change the manifestation of gender in their speech and dialect markers are similarly permanent and pervasive (although they are more easily modified than gender-related features). Sociolinguistic characteristics like g-dropping and vowel quality, on the other hand, change fluidly in everyday sociolinguistic interaction, as context shifts in formality and conversational partners.

It is tempting to conclude on the basis of these data that conscious perceptual categorization, therefore, effectively overlooks those aspects of speech which are under the least direct control of speakers. Features which cannot easily be changed affect other aspects of perception – such as speed of word recognition or degree of speech intelligibility – but they might not play much of a role in the social categorizations taking place constantly during normal linguistic interaction. Instead, listeners may focus in on aspects of a person's speech that they choose, in some sense, to employ – like the g-dropping or vowel quality features in our stimuli.

The results of this experiment should be interpreted with caution however: they reflect the perceptual similarity judgments of a small group of listeners hearing tokens of only a single word, *going*, in a limited number of sociolinguistic guises from a small set of speakers. Generalizing these results to a greater number of base words in parallel sociolinguistic variations would lend a great deal of support to the conclusion that these more fluid, stylistic variables are more salient in listeners' perception. Replication with a larger body of listeners from diverse language background would further support the generality of the result. And in the future, listeners' perception of other fluid sociolinguistic variables, such as r-full versus r-less pronunciations, should be tested to probe the generality of the conclusion that these sorts of variables seem to "trump" more static factors like talker gender in perception.

Finally, there is a risk that speakers who were recorded during collection of the materials for this study were unconsciously emphasizing the speech characteristics that were most unusual for them, specifically the g-dropping and changes in vowel quality. To probe the potential effects of this unintentional speaker bias, materials taken from more naturally produced speech should be used in future work as well.

## 5. Conclusions

The results of this investigation into listeners' use of sociolinguistic variables in perception are suggestive of a strong effect of fluid speech characteristics in categorization and similarity judgment. With a limited number of talkers and tokens from only a single word of English, listeners weighed features such as g-dropping and vowel quality above talker dialect and gender in their judgment of perceptual similarity. Further studies are needed to test the generality of the privileged status of fluid features in perception, but if it holds true that listeners pay the most attention to the very factors that variationist studies of production have highlighted again and again as social markers, then perhaps the void between what is currently known about the role of sociolinguistic variation in production and its use in perception is not as wide as it might seem.

## 6. References


Campbell-Kibler, K. (2006). *Listener perceptions of sociolinguistic variables: The case of (ING)*. Unpublished doctoral thesis, Stanford University.

Campbell-Kibler, K. (2007). Accent, (ING), and the social logical of listener perceptions. *American Speech. 82*, 1. 32-64.

Campbell-Kibler, K. (2009). The nature of sociolinguistic perception. *Language Variation and Change, 21*, 1. 135-156.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In Luce, R. D., Bush, R. R. & Galanter, E. (Eds.) *Handbook of Mathematical Psychology*. New York, NY: Wiley.

Clopper, C. G., & Pisoni, D. B. (2004a). Effects of talker variability on perceptual learning of dialects. *Language and Speech, 47*, 207-239.

Clopper, C. G., & Pisoni, D. B. (2004b). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics, 32,* 111-140.

Clopper, C. G., & Pisoni, D. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics, 35*, 421-438.

de Wolf, G.D. (1988). On phonological variability in Canadian English in Ottawa and Vancouver. *Journal of the International Phonetic Association, 18*, 2. 110-124.

Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics, 12*, 4. 453-476.

Fischer, J.L. (1958). Social influences in the choice of a linguistic variant. *Word, 14*. 47-56.

Johnson, K. A., (2005). Speaker normalization in speech perception. In: Pisoni, D. B., & Remez, R. E. (Eds.) *Handbook of Speech Perception*. Malden, MA: Blackwell Publishing, Ltd.

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

Kruskal, J. B., & Wish, M. (1978). *Multidimensional Scaling*. London: Sage Publications.

Labov, W. (1966). *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., & Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *Journal of the Acoustical Society of America, 59*, 2, 675-678.

Munson, B., Jefferson, S. V., & MacDonald, E. C. (2006). The influence of perceived sexual orientation on fricative identification. *Journal of the Acoustical Society of America, 119*, 4, 2427-2437.

Romney, A. K., Brewer, D. D., & Batachelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science, 4*, 1. 28-34.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika, 27*, 3, 210-245.

Tamati, T. (in press). Perceptual similarity of unfamiliar dialects: Some preliminary findings. *Research on Spoken Language Processing: Progress Report No. 29*. Bloomington, IN: Cascadilla Press.

Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization*. Hillsdale, NJ: Erlbaum.

Wardhaugh, R. (2006). *An Introduction to Sociolinguistics* ($5^{th}$ ed.). Malden, MA: Blackwell Publishing.

Woods, H.B. (1979). *A socio-dialectology of the English spoken in Ottawa: A study of sociological and stylistic variation in Canadian English.* Unpublished doctoral thesis, University of British Columbia.