



University of Pennsylvania Working Papers in Linguistics

Volume 7

Issue 3 *Papers from NWAV 29*

Article 18

1-1-2001

Chipping Away at the Perception/Production Interface

Nancy Niedzielski
niedz@rice.edu

Chipping Away at the Perception/Production Interface

Chipping Away at the Perception/Production Interface

Nancy Niedzielski

1 Introduction

One of the most curious aspects of language variation is the fact that there are some groups of speakers who, despite constant and copious amounts of acoustic evidence to the contrary, persist in their belief that they speak standard, "unaccented" American English. Several studies have shown, for instance, that speakers who reside in the cities most prominently involved in the Northern Cities Chain Shift (NCCS) (e.g., Detroit middle class whites) are just such a group (cf. Preston 1989, Frazer 1993 (and the articles therein), Niedzielski & Preston 1999). These speakers, whose vowel systems diverge in myriad ways from what one might consider the more standard vowel formant values of Peterson & Barney (1952), and who are surrounded by speakers whose vowels also diverge, often report that their speech community is the one of the last bastions of Standard American English (SAE). A respondent from southeastern Michigan quoted in Niedzielski & Preston (1999:99) demonstrates this:

...I think what...K was talking about was the standard—if you have such a thing as called standard English other than textbook English, it would probably be the language that you're hearing right now. As you listen to the Midwestern.

Several respondents in this same data set made similar statements, often offering opinions regarding national news anchors and television personalities, who use "Midwest voices" (98) or Midwestern English. The linguistic security of the Michigan speaker is thus illustrated.

There are two possible hypotheses about why Northern Cities speakers continue to report that their vowels and the vowels of their cohorts are standard: 1) the chain-shifted vowels of these regions are now considered to *be* standard, so that these vowel formant values have replaced the earlier canonical vowel formant values of standard speakers; or 2) Northern Cities speakers do not perceive (at some level) these chain-shifted formant values in the speech of their fellow speech community members (but must perceive them at another level, since these are the vowels that they produce).

Previous work, which I have reported on elsewhere, provided evidence for the second hypothesis (cf. Niedzielski 1997, 1999a). In this paper, I

present data that suggests that chain-shifted vowels may be becoming more salient to speakers from Detroit (and its immediate environs), but that they are still not salient in the speech of those people that Detroiters feel belong to their speech community. I will also present an initial attempt at a model of speech perception that presents different perceptual levels to account for 1) the ability of Detroiters to perceive NCCS vowels in order to be able to produce them; 2) the inability to perceive these vowels if a speaker is thought to belong to a Detroiters' ingroup;¹ and 3) the ability to perceive these vowels if a speaker is thought to not belong to a Detroit speaker's ingroup.

2 The Influence of Social Factors on Perception

In the above-mentioned previous work, I summarize in detail the research that demonstrates that speech perception involves much more than the simple transformation of acoustic information into linguistic categories. Research shows that visual information plays an important part in speech perception (cf. McGurk and MacDonald 1976), and phonemic categories in a speaker's language play an important role (cf. work on categorical perception: Beddor and Strange 1982, etc.; work on vowel-space calibration: Ladefoged and Broadbent 1957, etc.; work on vowel discrimination and dialect: Willis 1972, Janson 1986, etc.).

In addition, social factors have been shown to influence speech perception. Strand and Johnson (1996) demonstrate that perceived gender has an effect on the perception of certain fricatives. They show that the presentation of a male versus female photograph as the producer of an utterance affects what fricatives subjects perceive, and suggest that speaker normalization (and thus speech perception) is "based on the perceived [gender] identity of the speaker" (25).

My research on perceived nationality in Detroit (Niedzielski 1997) provided evidence for the effect that an *a priori* category could have on the reported perception of vowel tokens in a speaker's speech: the labels "Michigan" and "Canadian," when applied to a sample of recorded speech, had an effect on which synthesized vowel tokens subjects offered as a best match to the speaker's speech. Subsequent research (Niedzielski 1999b) showed that the categories of 'real speech' or 'synthesized speech' seemed to have an effect on subjects' perceptions of certain fast speech phenomena, since subjects reported noticing such phenomena in recorded speech that

¹ That is, the group of people who, like them, use SAE.

they thought was synthesized, but not in speech that they thought was human.

Taken together, these studies demonstrate that several factors, including perceived identity of a speaker, effect the perception of speech, and have the potential to override acoustic information in the processing of a speech signal. They also offer the first piece to the puzzle of how speakers can use supposedly contradictory information about speakers in their own speech community to construct a "standard speaker" identity.

3 Detroiters' Perceptions of SAE Regions

There has been ample research showing that white middle class residents of southeastern Michigan believe that they are at ground-zero for Standard American English usage. The respondent quoted in section 1 above exemplifies the typical viewpoint of these residents—that their English represents the standard for American English. Additional research demonstrating such language attitudes of Michiganders is presented in, for example, Preston (1987) (which includes hand-drawn maps of subjects showing that Michigan, and only Michigan, receives their highest "correctness" ranking), and Niedzielski (1997) (which includes data showing that 28 out of 30 respondents offered Michigan as an example of where Standard English was found, with Great Britain being the only other location suggested).

Recent language attitudes work which I conducted in the summer of 2000 suggests, however, that such speakers may be becoming more aware of non-standard features in the speech of people who, though still perceived as outgroup members, reside in an increasingly closer area to themselves. I conducted a written survey of 30 residents of southeastern Michigan,² in which I asked questions regarding specific areas close to Detroit, and whether those language varieties differed from their own. I asked them whether or not they would expect to hear dialect differences in speakers from states surrounding Michigan, particularly Wisconsin and Minnesota, and to be specific about those differences.

Results of this study seemed to suggest that these residents are beginning to feel that speakers of non-standard varieties of English are creeping closer and closer. While 7 out the 30 respondents stated that there were no differences in the dialects of speakers from surrounding states (most often noting that these states were "accentless" or that there was no accent in the Midwest), 23 out of the 30 stated that there *was* a definite difference in the

² Specifically, from the cities of Ann Arbor, South Lyon, and several Detroit-area suburbs.

speech of speakers from the states surrounding Michigan: 9 out of these 23 respondents listed Ohio; 9 listed Wisconsin; 9 listed Minnesota; 3 listed Illinois (specifically, Chicago); 2 listed Indiana;³ 8 listed Canada; and 3 of these mentioned Windsor, Ontario specifically.⁴

Even more compelling is the fact that while several questions asked for information about specific states, 14 out of the 23 respondents who noted dialect differences offered some part of Michigan as example of such an area, with 12 mentioning the Upper Peninsula, and two noting that they expected differences in speakers from the northern part of the lower peninsula (specifically Houghton Lake and "rural towns north of [Ann Arbor]").

Specific features that respondents offer as examples of differences (with features that correspond to those offered in an earlier language attitudes study [Niedzielski 1996] **bolded**) include:

- (1) **use of the discourse marker *eh* (for Upper Peninsula, Canada)**
- (2) the use of /d/ for /θ/
- (3) **a marked production of /æ/ (which several respondents illustrated by stating "the vowel in *dad*" or "the pronunciation of *bag*")**
- (4) **a marked production of /o/ (most often illustrated by noting "the 'o' in *Minnesota*")**
- (5) a marked production of /s/
- (6) the use of *ya* for *you*
- (7) the use of *soda* for *pop*

In addition, respondents offered qualitative evaluations including: "relaxed," "lazy," "hockey voice," "drawn-out vowels," "Scandinavian-sounding," and statements such as "the letters are not pronounced properly," "words are over-pronounced" and "the grammar is bad/poor."

Several respondents also noted that the proximity of Minnesota and the Upper Peninsula to Canada led to speakers from those areas sounding Canadian (apparently ignoring the fact that only the Detroit River separates Detroiters from Canadians).

³ Several respondents listed more than one location.

⁴ Windsor is directly across the Detroit River from downtown Detroit; one crosses this international boundary via the Windsor tunnel under the river, or the Ambassador Bridge over it.

⁵ The respondent merely listed "the letter s" as illustrative of differences found in Minnesota.

The attitudes survey results presented here seem to suggest, then, that it is *not* that case that most speakers from southeastern Michigan believe that all Midwesterners (or even all Michiganders) speak SAE—most of the features that they offer as examples are clearly, for them, non-standard. Rather, this suggests that these residents believe that they are barely keeping the SAE walls from collapsing in, and that non-standard dialects are rapidly closing in on them. The number of people who comprise their ingroup of SAE speakers is relatively small, and the outgroup members numerous.

While it is impossible to state exactly what features the subjective comments refer to, descriptions of the more specific features suggest that at least some of them are found in the respondents' own dialect. For instance, the three phonological illustrations—/æ/-fronting (mentioned by five respondents as characteristic of Minnesota/Wisconsin speech), /o/-monophthongization (mentioned by four respondents as characteristic of Minnesota speech), and /θ/-stopping (mentioned by two respondents as characteristic of Minnesota and Upper Peninsula speech)—are all commonly found in Detroiters' varieties. In a quick acoustic analysis of one southeastern male Michigander's and one male Minnesotan's speech,⁶ I found that both speakers used /d/ in words like 'there' and 'then,' both speakers consistently produced /æ/ with an F2 of about 2250 Hz (considerably higher than the F2 value of 1750 Hz that Peterson and Barney give for this same vowel produced by a male speaker), and both speakers produced /o/ without an off-glide. Thus, at least some of the features that these Detroiters mentioned are likely found in their own dialects.

This suggests a second piece of our puzzle: though these speakers *do not* notice non-standard features in speakers from their ingroup, they *do* notice them in speakers from areas close to them. They do not filter out acoustic information from *all* Midwestern speakers—only those who belong to their own perceived speech community, which is presumably shrinking. In other words, they notice NCCS "non-standard" features, but not among members that they assign to their ingroup.

4 A Proposed Model of the Perception/Production Link

Two features that none of the respondents mentioned (and that none of the respondents in Niedzielski (1996) mention) are features that were in fact different in the Michigander's and Minnesotan's speech (Niedzielski 1996). In the quick acoustic analysis mentioned above, I found several examples of

⁶ This study will include many more residents from both places in the final version of this research.

words that illustrated a o/a merger in the Minnesotan's speech, and not in the Michigander's speech, and several examples of $/\text{e}/$ produced with no off-glide in the Minnesotan's speech, but with an off-glide in the Michigander's speech. While it is impossible to state that Michiganders do not notice these features (they may have simply not listed them on the survey forms or in the interviews), it is tempting to suggest that these features are at least not as salient to Detroiters as those that *were* reported in these two language attitudes surveys.

I therefore hypothesize here that the fact that Detroiters report the use (in others' dialects) of the features they produce themselves, but do not report the features that they do not produce, is significant. In other words, the fact that these features are produced by these Michigan speakers is a key to why these features are perceived, at least in others' dialects, and conversely, this may partially explain why speakers do not "perceive," yet are able to produce, certain phonological features

Young children acquiring the dialect of their speech community *must* be able to perceive the features of that dialect—there would be no other way for them to reproduce these features. At this particular developmental stage, there may be a closer match between what a person perceives, what he or she would report perceiving, and what he or she produces. A child able to perform a matching task similar to the one given to respondents in Niedzielski (1997) may not be influenced by the nationality labels as the adults were, for a number of reasons.

First, and most obvious, is that children may not have acquired the social categories that adults have: "Canadian" speaker, for instance, may not be a relevant category for a child (although there is evidence that some social (or biological?) categories such as "male" and "female" speaker are relevant even in infants). Second, and related to the first, a child may not have had any exposure to speakers from these different social groups.

Third, and I suggest that this may be most important, the child may have no social *reason* to "misperceive" his or her own dialect. Part of the linguistic development of a speaker must include the acquisition of value judgments about certain language variants, which is something that is acquired later than features of the language variants themselves.

Meyerhoff (2001) illustrates how such a process might operate by appealing to the social-psychological motivations of the speaker. She suggests that two of the main motivations for language variation are 1) accruing social capital and 2) minimizing risk. To take the former case, some language varieties index a group's "access to social or economic capital (62)," or are "metaphorically associated with the trappings of this access," thus whether they are actual or virtual signs of access is irrelevant. It would seem, then,

that giving up the belief that one's variety is a symbol of such trappings is not a simple or fast process, and it might take an overwhelming amount of evidence to the contrary for one to give up such a belief. At this point in sociolinguistic history, there is very little in, for instance, popular culture that would cause a Detroiters to feel that his or her variety offers anything but direct access to capital, so there is not reason to pay attention to any evidence to the contrary.

In contrast, a strong motivation for the *avoidance* of producing some variants involves "distancing yourself from behaviors and groups of people associated with the peripheries (63)." There is much in popular culture to tell speakers the cost of using nonstandard dialects, and so there must be a strong motivation to hold onto the belief that one is in no way a part of "the periphery," until there is compelling evidence to the contrary.

Thus, while Meyerhoff attributes these motivations to the *production* of certain variants (or, in some cases, their avoidance), I hypothesize that this, too, is a strong motivation for perception. Unless speakers are forced to give up the belief that they can accrue social capital and minimize risk by using their language variety, then they will continue to believe this despite contradictory acoustic evidence—the cost to believe otherwise is too great.

Additionally, tasks such as those that I required of my respondents automatically highlight ingroup-outgroup distinctions. If such contradictory acoustic evidence is made salient—by, for instance, the media, or a researcher overtly telling a speaker to evaluate features of a dialect, then the ingroup—outgroup becomes particularly important. When the label that is applied to the test tape speaker corresponds to a respondent's ingroup, I hypothesize that the speaker is evaluated as a member of the respondent's ingroup, but on an intergroup level, and that all of the beliefs that the respondent has about his or her own speech are "perceived" in the speaker's variety. On the other hand, when the label corresponds to what the respondent assigns to an outgroup category, that speaker is evaluated on an interindividual level—the respondent asks, perhaps, how specifically does this speaker's variety differ from mine?—and thus greater attention is paid to specific features in that speaker's speech.

5 What Speech Perception Research Offers

There are several aspects of basic models of speech perception that can be appealed to in order to explain the processes that would need to operate in order for such an analysis to be valid.

First, there are several theories that might provide an explanation for the fact that what Michiganders report perceiving corresponds almost exactly to

vowel tokens that are often offered as typical of "general American vowels" (such as those found in introductory linguistics or phonetics texts). Johnson, Flemming, and Wright (1993) proposes the notion of a vowel "hyperspace" to explain why, when subjects are asked to synthesize a vowel that matches one heard on a test tape, they consistently offer a vowel that is more peripheral than the one the speaker actually produced. They contend that what people perceive is closer to "hyperarticulated" versions of vowels (that is, tokens on the periphery of the vowel space), rather than the more centralized versions that people actually produce. They hypothesized that subjects matched tokens to speech based on "reference to a representation of that sound in memory (523)," and that these representations were "hyperarticulated." I contend that that is what people do who are from areas where the "standard" purportedly is used. These speakers, when given the task of choosing tokens that match vowels found in a speaker thought to be in their ingroup (that is, fellow "standard" speakers), refer to a representation of a more hyperarticulated version of the vowel—one that is not centralized, raised, backed, etc., but rather a token that would be a part of a hyperarticulated system.

Second, speech perception models offer processes that can account for the different degrees of attention that I propose speakers give to ingroup versus outgroup speakers. I contend that respondents match a token to an idealized representation in the case of a perceived ingroup member, but pay closer attention to actual features in the speech of perceived outgroup members. This accounts for the fact that respondents in Niedzielski (1997) were much more accurate in choosing tokens that match the test tape speaker's vowels if they thought (s)he was an outgroup member, but not very accurate if they thought (s)he was an ingroup member, for instance. Nusbaum and Schwab (1986) make reference to the parallel of active and passive processes that occur in speech perception, claiming that passive processes should be "automatic," and active processes should be "cognitive." The automaticity of passive processes means that passive processing should be invariant, or at least should represent "a lack of flexibility in responding to a stimulant (123)." If a respondent is appealing to the fixed representation of a vowel token, for instance, then the "perception" should be relatively more invariant. However, active processes require more cognitive effort, and thus more attention. I contend that in allowing respondents to believe that a speaker is a member, not of one's ingroup, but rather an outgroup member, I forced respondents to use more active processing strategies in evaluating their speech, and this led to greater accuracy in token matching.

Finally, I contend that the active/passive process distinction may account for the preliminary finding of speakers identifying features that they

share with "outgroup" dialects more often than features that they do *not* share. I suggest that at some point during the acquisition of one's native language variety, the perception of that ingroup variety must involve active processes, even if one is acquiring a dialect that most people believe to be "standard." I hypothesize that those active processes may increase the salience of non-standard features in the ingroup variety, so that at a later developmental stage, when speakers no longer include these features as part of the their ingroup's phonological system, they are still able to identify these features, though now as "outgroup" features.

6 Conclusion

The phenomena described here illustrate the very real "working" nature of this paper, in that they produce more questions than answers. Is the hyperspace hypothesis adequate for explaining where our ideas of SAE come from? Is the ingroup/outgroup distinction really what is behind the very different results found when respondents are given different labels for speakers? Does the fact that non-standard features speakers share are more salient have any bearing on the production/perception interface? While work in social psychology and speech perception can offer insight into certain observed behavior, without overtly testing the models for their applications to expectations and dialect perception, it remains speculation.

It nonetheless seems valuable to continue research on dialect perception by appealing to work in speech perception and social psychology, paying particular attention to the processes found in standard models in those fields. The next round of research in this area of dialect perception will do just that.

References

- Beddor, Pamela and Winifred Strange. 1982. Cross-language study of the oral-nasal distinction. *JASA* 71:6. 975-989.
- Frazer, Timothy C. 1993. *Heartland English: Variation and Transition in the American Midwest*. Tuscaloosa: University of Alabama Press.
- Janson, T. 1986. Sound change in perception: an experiment. In Ohala, John and Jeri Jaeger, (eds). *Experimental Phonology*. Orlando: Academic.
- Johnson, Keith, Edward Flemming and Richard Wright. 1993. The hyperspace effect: phonetic targets are hyperarticulated. *Language* 69:3. 505-528.
- Ladefoged, Peter and D. E. Broadbent. 1957. Information conveyed by vowels. *JASA* 29:1. 98-104.
- Meyerhoff, Miriam. 2001. Dynamics of differentiation: on social psychology and cases of language variation. In Coupland, Nikolas, Christopher Candlin and Sri-

- kant Sarangi (eds). *Sociolinguistics and Social Theory*. London: Longman. 61-87.
- Niedzielski, Nancy. 1999a. The effect of social information on the perception of socio-linguistic variables. *Journal of language and social psychology* 18:1. 62-85.
- Niedzielski, Nancy. 1999b. Speech perception and speech synthesis. Paper presented at LSA. Los Angeles, CA.
- Niedzielski, Nancy. 1997. *The effect of social information on the phonetic perception of sociolinguistic variables*. Unpublished doctoral dissertation. University of California, Santa Barbara.
- Niedzielski, Nancy. 1996. Acoustic analysis and language attitudes in Detroit. In Meyerhoff, Miriam (ed.) *(N)Waves and means: a selection of papers from N-WAVE 24*. Philadelphia: U. Penn Working Papers in Linguistics.
- Niedzielski, Nancy and Dennis Preston. 1999. *Folk Linguistics*. Berlin: Mouton
- Nusbaum, Howard and Eileen Schwab. 1986. The role of attentive and active processing in speech production. In Schwab, Eileen and Howard Nusbaum (eds.). *Pattern recognition by humans and machines: speech perception*. Orlando: Academic.
- Peterson, Gordon and Harold Barney. 1952. Control methods used in the study of vowels. *JASA* 24: 2175-2184.
- Preston, Dennis. 1989. *Perceptual dialectology*. Dordrecht: Foris Stand, Elizabeth and Keith Johnson. 1996. Gradient and visual speaker normalization in the perception of fricatives. In Gibbon, Dafydd (ed). *Natural language processing and speech technology*. Berlin: Mouton.
- Willis, Clodius. 1972. Perception of vowel phonemes in Fort Erie, Ontario, Canada and Buffalo, New York: an application of synthetic vowel categorization tests to dialectology. *Journal of speech and hearing research* 15:246-255.

Rice University Department of Linguistics
Houston, Texas 77251-1892
niedz@rice.edu

Why You Can't Do a VARBRUL Study of Quotatives And What Such a Study Can Show Us

John Victor Singler

1 Introduction

The twentieth century saw the introduction of three quotatives into American English.¹ First there was *go*, whose appearance appears to date at least as far back as the 1940's and 1950's, according to the recollection of those who were teenagers then. After *go* came *(be) like*, first noted in Butters (1982). In some parts of the United States, notably California, *(be) all* has now followed—and to some degree supplanted—*(be) like*. Three consecutive quotatives in the course of a narrative by a New York City female college student

¹The data for this study were provided by Samantha Abrahamsohn, Nicole Abrams, John Allison, Bryan Anderson, Chris Anderson, Yuliya Angelovskaya, Heather Barlin, Marisa Bassi, Sharmila Basu, Frada Berenshteyn, Amanda Bernstein, Dana Bivacca, Ashley Blackwood, Melissa Boltax, Jill Boncek, Abigail Braithwaite, Cristina Buccola, Steve Carver, Natalie Charles, Pete Chatmon, Kimberly Cook, Laurie Crist, Karen Cummo, Kelly Cunningham, Niels Dachler, Lila Damavandi, Roslyn Dames, Larissa Davidov, Samantha DeCerce, Jay Diaz, Danielle DiTalia, Florence Emerole, Rachael Evans, Sawyer Fischer, Andrew Fleming, Michael Foster, Jessica Geminder, Greg Goldberg, Suzann Goldberg, Lauren Goodman, Amanda Hamann, Shari Headley, Dana Horoszewski, Francis Hult, Noriko Ishibashi, Aiko Ishikawa, Maria Isidorou, Orly Jalowski, Clement Joseph, Ericka Joseph, Jennifer Kaplan, Katherine Kearney, Allegra Ketchum, Christine Kim, Heedal Kim, Sun Kim, Jennifer Kizielewicz, Stavroula Kolitsopoulos, Sarah Kramer, Tatum Kutzer, Andria Kyriakides, Michael Landis, Abby Leap, Eugene Lee, Joann Lee, Hao Hong Lin, Stephanie Lloyd, Nikiasha London, Sofia Lopez, Shelby Lovecchio, Michael Makarius, Tom Mandulak, Christopher Martino, Yvonne Martin, Michael McCarthy, Ruth McKee, Molly McIver, Shin Mitsugi, Yuliya Motuz, Christin Muller, June Nakajima, Ndeye Ndiaye, Laina Niciu, Shunsuke Nozawa, Helen Papaioannou, Joo-young Park, Eleni Passias, Benjamin Perriello, Jackie Pinkas, Karen Polanco, Amy Posner, Christopher Potts, Chris Poultney, Elizabeth Pratt, Daniela Raik, Peter Raucci, Alex Raytburg, Tiffany Robertson, Eleni Sakellis, Jesse Salazar, Christina Schlegel, Lina Shah, Linda Shipley, Sara Simonetti, Juyoun Song, Brian Stillman, Brian Struck, Natalie Toussaint, Suzanna Touzet, Adrienne Tremblay, Tony Tsai, Polina Veksler, Kathryn Venverloh, Jon Villanueva, Polina Veksler, Masaki Yamataga, and Terra York. I am grateful to them and to Maryam Bakht-Rofheart, Richard Cameron, David Heap, Erez Levon, and Linda Susman for helpful comments.