CrossMark click for updates

Effects of phonetically-cued talker variation on semantic encoding

Meghan Sumner[a)]
Department of Linguistics, Stanford University, Margaret Jacks Hall, Building 460, Stanford University, Stanford, California 94301-2150
sumner@stanford.edu

Reiko Kataoka
Department of Linguistics and Language Development, San José State University, One Washington Square, San José, California 95192–0093
reiko.kataoka@sjsu.edu

Abstract: This study reports equivalence in recognition for variable productions of spoken words that differ greatly in frequency. General American (GA) listeners participated in either a semantic priming or a false-memory task, each with three talkers with different accents: GA, New York City (NYC), and Southern Standard British English (BE). GA/BE induced strong semantic priming and low false recall rates. NYC induced no semantic priming but high false recall rates. These results challenge current theory and illuminate encoding-based differences sensitive to phonetically-cued talker variation. The findings highlight the central role of phonetic variation in the spoken word recognition process.

© 2013 Acoustical Society of America
PACS numbers: 43.71.Es, 43.71.Sy [AC]
Date Received: July 11, 2013 **Date Accepted:** October 6, 2013

1. Introduction

Listeners quickly and adeptly understand spoken words despite massive variation in the speech signal. Episodic theories of lexical access accommodate talker-specific variation in speech perception (e.g., Goldinger, 1998). These theories suggest that individual linguistic events are encoded, with frequent events forming dense clusters of episodes. These clusters facilitate activation of acoustically similar productions compared to less dense clusters (see Johnson, 2006). The bulk of work examining variation has used form-based tasks, potentially leading to predictions that may or may not be supported when considering a deeper level of encoding. Semantic encoding measures provide a way to investigate effects of variation at this deeper level, providing breadth that may lead to improved explanations of spoken language understanding.

Recent work has shed light on issues surrounding frequency-based approaches to speech perception. Sumner and Samuel (2009) found that General American (GA) listeners new to the New York City (NYC) area recognized semantically related targets (e.g., *thin*) faster after critical primes (e.g., *slender*) than after unrelated control words, but only when primes were produced by a rhotic-GA talker (e.g., *slend-er*). No such facilitation was found when primes were produced by a non-rhotic-NYC talker (e.g., *slend-uh*). Importantly, *both* types of primes facilitated target recognition equally for rhotic- *and* non-rhotic participants living in the NYC area. This is a case of recognition *equivalence.* Because semantic activation is delayed when form-based processing is slowed (van Orden and Goldinger, 1994), we would not expect less frequent productions to yield strong semantic priming. This effect appears to be independent of quantitative frequency. As a further wrinkle, listeners have better memory for infrequent,

[a)]Author to whom correspondence should be addressed.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 171.65.238.130 On: Thu, 20 Feb 2014 22:21:49

perceived standard forms, in that such forms are remembered better in long-term form-based tasks independent of listener accent (Sumner and Samuel, 2005, 2009). These data do not reflect abstraction; they reflect detailed memory for infrequent but idealized forms of words.[1]

This study serves two purposes. First, we investigate the recognition of words produced by talkers with different non-GA accents as a direct test of the effect of frequency of exposure in spoken word recognition. Differences in frequency may certainly exist across different accents of English. But, by any standard measure of global frequency, words uttered in GA accents should be more frequent for GA participants than those uttered in various non-GA accents. Second, we explore effects of phonetically-cued talker information on semantic encoding using a convergent task methodology. We investigate the recognition of spoken words across talkers with different accents in the semantic priming paradigm (Experiment 1a) and the false memory paradigm (Experiment 1b). Our talkers' accents are broadly classifiable from variable phonetic cues [GA, Southern Standard British English (BE), NYC].[2] Our non-GA talkers differ in *perceived* standard variants—a non-rhotic vowel may be standard or nonstandard depending on accompanying phonetic variation (e.g., a BE voice vs a NYC voice). As phonetic variation cues listeners to talker information, we used two accents that prompt different qualitative perceptions: BE is perceived broadly to be the standard accent of English and is accorded high prestige (e.g., Coupland and Bishop, 2007) and non-rhotic NYC is perceived negatively by GA speakers (Preston, 2002). In experiment 1a, we investigate the semantic priming of visual targets by related and unrelated primes produced by our three talkers for GA listeners. We ask whether final *-er* words spoken by our GA and non-GA talkers facilitate target recognition. Based on past work, we expect GA primes to facilitate responses to semantically related targets and NYC primes to preclude facilitation. From what we know about episodic lexical access, responses to targets following non-GA primes should pattern together and differ from those to targets following GA primes. Any split between the two non-GA accents would not easily be captured in a purely quantitative approach to spoken word recognition. One drawback of the semantic priming paradigm is the potential for a null effect, leaving one to ponder the cause of the lack of target facilitation. As mentioned, slowed form-access delays semantic activation, so a paradigm that eliminates the potential for a null effect *and* provides ample time for semantic activation is desirable. In experiment 1b, we address these concerns via the false memory paradigm. In this paradigm, listeners are presented with a list of semantic associates (*bed, rest, awake,* etc.) that are related to an unstudied lure (*sleep*). A false memory results when participants erroneously recall the lure as a studied word. Accuracy in recalling studied words and unstudied lures is measured. Two main accounts of false memories exist. First, the associative activation of semantically related words leads to false memories (Roediger and McDermott, 1995). Second, words are encoded with different strengths of verbatim or gist memory traces (see Brainerd *et al.*, 2008). In other words, listeners encode exactly what was said (verbatim) or the general idea of what was said (gist). When encoding the general idea, false memories result as the semantic set becomes and remains active. When encoding exactly what was said, listeners have better memory for particular words, limiting semantic spread and aiding in recall. This second perspective explains differences in false recall across populations and listening conditions (e.g., children have better verbatim memory than adults who typically extract the general meaning of an utterance; see Brainerd *et al.*, 2008; Otgaar *et al.*, 2012). We are again interested in response similarities and differences across (GA and non-GA) talkers.

2. Talker-based effects in paradigms sensitive to semantic encoding

2.1 Methods

2.1.1 Participants

Seventy-four monolingual English-speaking undergraduates participated. Sixty participants were assessed via a post-experiment questionnaire to be GA speakers, born and raised in rhotic regions of the U.S., and to be children of monolingual

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 171.65.238.130 On: Thu, 20 Feb 2014 22:21:49

English-speaking parents. Thirty-four participated in experiment 1a and 26 participated in experiment 1b.

2.1.2 Stimuli

Speakers from England and New York were recruited via fliers that made no mention of accent. Four talkers were chosen as our non-GA talkers (two non-rhotic NYC, two non-rhotic BE) along with the first author as our GA talker. All talkers were female in their mid-thirties. The non-GA talkers had lived in California for less than 3 months at the time of recording. One BE and one NYC talker were used in experiment 1a; the other two were used in experiment 1b. Different talkers across the experiments minimized effects of a specific talker. All talkers produced the stimuli three times at a comfortable speaking rate in a sound-attenuated booth. Tokens were chosen based on similarity in duration and were scaled to have the same peak amplitude. To assess intelligibility in the clear, three counterbalanced lists of the stimuli were presented to native GA listeners who entered each word on a keyboard. All items were identified at ceiling with error rates of 2.06% (GA), 2.01% (BE), and 2.09% (NY). To assess the social perceptions of each voice, we played subsets of these lists (30 words per single-talker list) to a separate set of 60 listeners who rated each voice on a scale of 1-6 for various traits (trait/GA, BE, NY: Clear/5.25, 4.90, 4.85; elegant/3.10, 4.57, 2.50; familiar/5.70, 4.81, 4.85; harsh/2.4, 3.04, 4.25; normal/4.65, 4.43, 3.95; proper/4.10, 5.14, 3.45; standard/4.7, 4.58, 3.7; tough/2.65, 3.19, 4.60; understandable/5.61, 5.14, 5.15; uptight/2.5, 4.52, 3.7).

2.1.3 Post-experimental questionnaire

A post-experimental questionnaire was used to collect demographic information (e.g., previous locations lived, language background for participants and their families). We also collected accent-familiarity ratings (1 = unfamiliar; 7 = extremely familiar). The mean familiarity rating for our participants was 5.92 (GA), 4.38 (BE), and 4.45 (NYC). We also elicited three free-response descriptors per talker for each participant. These responses guided the choice of traits in the social ratings described earlier.

3. Experiment 1a: Semantic priming

3.1 Materials and design

Eighty-four critical primes were used in this study. Primes were chosen from an online semantic-association task in which 250 two- or three-syllable words ending with *-er* were presented to participants who provided the first word that came to mind for each of 125 randomized probes. Approximately 160 responses were collected per probe. The 84 *-er* ending words with the highest number of shared responses were used along with their corresponding targets (mean shared responses = 43%, minimum = 30%). Target words were presented in lower case, Times New Roman font, size 86 pt. Filler targets (252: 84 real, 168 pseudowords) were preceded by real word primes. Six counterbalanced lists were created so that each critical target was preceded by a related and unrelated prime across lists. Each list was counterbalanced for talker and included 42 related pairs (14 GA, 14 BE, and 14 NYC primes) and 42 unrelated pairs (14 GA, 14 BE, and 14 NYC primes). One-third of the filler trials were presented in each accent.

3.2 Procedure

Each participant was tested on one randomized list. A trial consisted of an auditory prime, a 100 ms ISI, followed by a visual target. A short ISI was used to allay concerns about the fading of potential effects of variation over a longer period and as an extreme timing comparison to the false-memory paradigm. Participants made lexical decisions to target words as quickly and accurately as possible. If no response was made within 3 s, a new pair was presented. If a response was made, the next trial began after 1 s. Participants were informed that the primes were spoken by three female speakers with different accents.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 171.65.238.130 On: Thu, 20 Feb 2014 22:21:49

3.3 Results and discussion

No differences in error rates were found, so analyses focused on reaction times (RTs) to correctly identified targets. RTs slower than 2.5 standard deviations from the each participant's mean and faster than 300 ms were excluded (2.3% of the data). Mean RTs to related and unrelated targets are provided in Table 1. A two-way (Talker × Relatedness) repeated measures analysis of variance (ANOVA) revealed a main effect of Relatedness [$F1(1,33) = 23.16$, $p < 0.001$; $F2(1,82) = 6.87$, $p < 0.05$] showing that targets preceded by semantically related primes were recognized faster than the same targets preceded by unrelated primes. No main effect of Talker was found [$F1(2,32) = 1.756$, $p = 0.189$; $F2(2,81) = 1.003$, $p = 0.371$], suggesting that listeners required the same amount of time to respond correctly to words independent of prime voice.

A Talker-by-Relatedness interaction was present by subject and item [$F1(2,32) = 3.781$, $p < 0.05$; $F2(2,82) = 3.136$, $p < 0.05$], suggesting that the priming effect was not uniform across conditions. Planned comparisons were performed on RTs and priming scores. Related targets were recognized faster than unrelated targets when preceded by GA and BE primes but not when preceded by NYC primes [GA: $F1(1,33) = 8.919$, $p < 0.01$, $F2(1,83) = 4.753$, $p < 0.05$; BE: $F1(1,33) = 7.482$, $p < 0.01$, $F2(1,83) = 4.163$, $p < 0.05$; NYC: $F1(1,33) = 1.016$, $p = 0.3208$, $F2(1,83) < 1$]. GA and BE primes resulted in a comparable degree of priming [$t(33) = 0.428$, $p = 0.671$], and both GA and BE primes resulted in greater priming than the NYC primes [GA vs NYC: $t(33) = 2.279$, $p < 0.05$; BE vs NYC: $t(33) = 2.334$, $p < 0.05$]. GA primes facilitate recognition to semantically related targets, but non-rhotic NYC primes do not. This now-replicated null effect should not be disregarded (as different talkers and paradigms were used), but we must consider it with caution. Critically, BE primes resulted in facilitation *on par* with GA primes. Across systematic variants (rhotic–non-rhotic) and talker accents (GA vs BE), we again find recognition equivalence. It is unlikely that lexical access is slowed for the BE talker as the short ISI should have revealed the effect of delayed access and reduced facilitation.

4. Experiment 1b: False memories across talkers

4.1 Materials and design

Eighteen semantic lists [15 from Roediger and McDermott (1995), three from Stadler *et al.* (1999)], consisting of a lure (e.g., *sleep*) and its 15 semantic associates (e.g., *bed, rest, awake*) were used. Six of the 18 study lists were presented in each accent. Three counterbalanced lists were created so all study lists were presented in each accent. List order was randomized. Arithmetic problem sets were created and used after the recall of each list.

4.2 Procedure

Participants were tested individually or in groups of two to three in a sound-attenuated booth. Associates were presented to participants ordered from most-related to least-related to lure. Participants then were instructed (per the standard) to type the words they heard beginning with the most recently heard items and then to freely recall of as

Table 1. Mean reaction times (ms) and error rates (in parenthesis in percentages) for semantically related and unrelated trials by prime condition.

Prime condition	Related	Unrelated	Difference
GA	508.0 (2.7)	531.6 (5.1)	−23.6*
BE	503.4 (2.3)	521.6 (5.4)	−18.2*
NYC	525.6 (3.4)	524.1 (5.2)	$1.5^{n.s.}$
Total	512.33 (2.8)	525.77 (5.23)	−13.43

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 171.65.238.130 On: Thu, 20 Feb 2014 22:21:49

many words as they remembered within 30 s. Participants then solved math problems for 45 s. A new set was presented after 45 additional seconds.

4.3 Results and discussion

We measured the proportion of presented words (veridical recall) and lures (false recall) recalled by our participants. The mean veridical and false recall rates by accent are provided in Fig. 1. A two-way [Talker (GA, BE, NYC) × Recall Type (veridical, false)] repeated measures ANOVA was conducted on the proportion of words recalled. Main effects of Talker [$F(2,50) = 3.909$, $p < 0.05$] and Recall Type [$F(1,25) = 32.826$, $p < 0.001$] and an Accent-by-Recall type interaction were found [$F(2,50) = 4.855$, $p < 0.05$]. Veridical recall rates were nearly identical across accents, speaking to the general intelligibility of our stimuli. Listeners falsely recalled more lures for NYC sets than for GA or BE sets [GA-NYC: $t(25) = -2.440$, $p < 0.05$; BE-NYC: $t(25) = 2.562$, $p < 0.05$]. Critically, listeners were equally good at *suppressing* lures for BE and GA sets [GA-BE: $t(25) = 0.102$, $p = 0.971$]. We found a consistent GA, BE - NYC asymmetry in both experiments that we address in the following section.

5. General discussion

We investigated the effects of phonetic variation on semantic encoding in the recognition and recall of spoken words. First and foremost, any account that depends on raw frequency of exposure will prove inadequate in accounting for these and other data. Across tasks, GA listeners recognize and recall spoken words uttered by a GA speaker and those uttered by a BE speaker with striking similarity. While we may yield to potential differences in exposure rates when comparing behavior to the BE and the NYC stimuli, that approach seems less fruitful when comparing behavior to the GA and BE stimuli for our listener population.

We see two possible and compatible accounts of these data. One possibility is that frequency, while important, may not be the best predictor explaining how listeners internalize linguistic events. Hearing a word, like *pretty* produced by a single speaker with a medial *tap,* 100 times in a default context (e.g., informal, casual speech) may indeed lead to a robust cluster of episodes. But hearing the same word uttered with a medial [t] in a socially salient context (e.g., teaching a child how to spell, where [t] co-

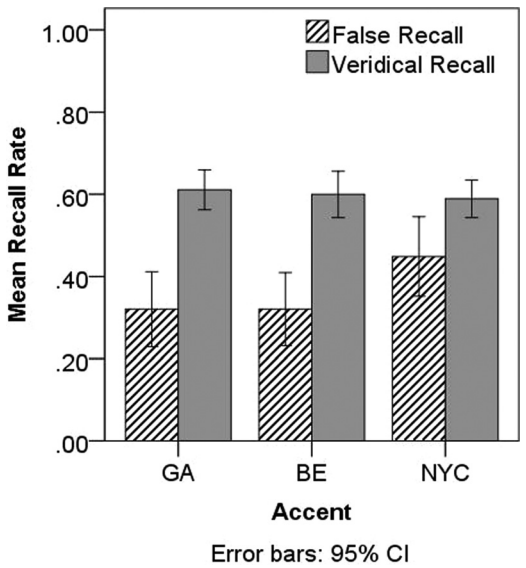

Fig. 1. Mean proportion of presented words correctly recalled (gray bars) and falsely recalled lures (striped bars) by talker.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 171.65.238.130 On: Thu, 20 Feb 2014 22:21:49

varies with clear speech in a context that specifically calls attention to the sounds of the words) may lead to stronger encoding (via increased attention). This *social weighting* may result in equally robust representations for variant pronunciations, despite differences in global frequency rates (see Sumner, 2013). This representational approach would suggest that the BE forms are encoded with greater detail than NYC forms. Equivalence between BE and GA that *appear* to be the same may result from the influence of a dense episodic cluster (GA) and a less dense, but more robust cluster (BE). While we believe this is likely, a representational approach alone does not account for the data as a whole. What we need to explain is not only how BE prompts equivalence to GA, but why NYC prompts a null effect in the semantic priming paradigm but robust spreading activation in the false memory paradigm.

A handful of studies have found that semantic priming is heavily modulated by attention (Smith *et al.*, 2001); semantic priming decreases as attention decreases. This effect is due specifically to attention allocated to primes (Otsuka and Kawaguchi, 2007). This finding links nicely back to recent explanations of false memories. The encoding of words with verbatim or gist traces is also modulated by attention. Specifically, gist encoding results from decreased attention; false recall rates are higher in divided attention conditions than full attention conditions (Otgaar *et al.*, 2012). Together, these would predict no priming in the semantic priming paradigm and high false recall rates in the false memory paradigm in reduced attention conditions. This pattern is exactly what we found for our NYC stimuli across two studies. Increased attention would predict an increase in semantic priming and a decrease in false recall rates—the attested pattern for our GA and BE conditions. Understanding why these cross-talker differences exist is a valuable endeavor but is beyond the scope of the paper. Speculatively, as effects of phonetically-cued talker information are much more pervasive than thought, it is likely that this information influences listener behavior. In this case, social differences cued by phonetic variation would modulate attention-based encoding.

We have suggested that episodic theories must be altered to accommodate recognition equivalence. In our view, this equivalence stems from access to episodic clusters that are dense with weakly encoded episodes (GA) and sparse with strongly encoded episodes (BE). In addition, we have found that phonetically-cued talker variation affects semantic encoding, as well as lexical access. This illuminates the fact that speech is a multi-faceted information source and suggests this information is highly influential in spoken language understanding.

Acknowledgments

We would like to thank Seung Kyung Kim, Ed King, and Kevin McGowan for helpful comments and discussion. Thanks also to Benjamin Lokshin for his help with data collection. This material is based upon work supported by the National Science Foundation Grant BCS-0720054 to M.S. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF.

References and links

[1]Were these effects due to abstraction, we would expect a high rate of false alarms for primes presented with non-standard variants. Instead listeners had highly detailed memory for words with the infrequent but ideal forms.

[2]NYC and BE differ from GA in a number of ways. The vowels for each are different from GA as documented for NYC (Gordon, 2004) and BE (Roach, 2004). NYC has a short-A split, lacking in GA (similar, but not identical, to BE), a contrast between [u] and [ru], and unmerged vowels before *r* yielding a two- or three-way distinction in the triplet *merry, Mary,* and *marry.* BE has [j] after coronals (optional after s, l), the vowel [ɑ] instead of [æ] in words like *bath* (though variable), and no *father-bother* merger (a three-way back vowel split for some). Other distinctions exist. There is sufficient reason to expect that obviously accented speakers of

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 171.65.238.130 On: Thu, 20 Feb 2014 22:21:49

each accent present listeners with considerable *less familiar* variation compared to native GA accents. There is little reason to expect *a priori* that one accent will be more or less difficult to understand than another.

Brainerd, C. J., Stein, L. M., Silveira, R. A., Rohenkohl, G., and Reyna, V. F. (**2008**). "How does negative emotion cause false memories?," Psychol. Sci. **19**, 919–925.
Coupland, N., and Bishop, H. (**2007**). "Ideologised values for British accents," J. Sociolinguist. **11**, 74–93.
Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.
Gordon, M. (**2004**). "New York, Philadelphia and other northern cities," in *A Handbook of Varieties of English. Phonology*, edited by E. W. Schneider (Mouton de Gruyter, Berlin), Vol. 1, pp. 282–299.
Johnson, K. (**2006**). "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," J. Phonetics **34**, 485–499.
Otgaar, H., Peters, M., and Howe, M. L. (**2012**). "Dividing attention lowers children's, but increases adults' false memories," J. Exp. Psychol. Learn. Mem. Cogn. **38**, 204–210.
Otsuka, S., and Kawaguchi, J. (**2007**). "Divided attention modulates semantic activation: Evidence from a nonletter-level prime task," Mem. Cogn. **35**, 2001–2011.
Preston, D. R. (**2002**). "Language with an attitude," in *The Handbook of Language Variation and Change*, edited by J. K. Chambers, P. Trudgill, and N. Schilling-Estes (Blackwell, Oxford), pp. 40–66.
Roach, P. (**2004**). "Illustration of British English: Received pronunciation," J. Int. Phonet. Assoc. **34**, 239–246.
Roediger, H. L., and McDermott, K. B. (**1995**). "Creating false memories: Remembering words not presented in lists," J. Exp. Psychol. Learn. Mem. Cogn. **24**, 803–814.
Smith, M., Bentin, S., and Spalek, T. (**2001**). "Attention constraints of semantic activation during visual word recognition," J. Exp. Psychol. Learn. Mem. Cogn. **27**, 1289–1298.
Stadler, M. A., Roediger, H. L., and McDermott, K. B. (**1999**). "Norms for word lists that create false memories," Mem. Cogn. **27**, 494–500.
Sumner, M. (**2013**). "A phonetic explanation of pronunciation variant effects," J. Acoust. Soc. Am. **133**, EL2–EL32.
Sumner, M., and Samuel, A. G. (**2005**). "Perception and representation of regular variation: The case of final-/t/," J. Mem. Lang. **52**, 322–338.
Sumner, M., and Samuel, A. G. (**2009**). "The role of experience in the processing of cross-dialectal variation," J. Mem. Lang. **60**, 487–501.
van Orden, G. C., and Goldinger, S. D. (**1994**). "Interdependence of form and function in cognitive systems explains perception of printed words," J. Exp. Psychol. Hum. Percept. Perform. **20**, 1269–1291.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 171.65.238.130 On: Thu, 20 Feb 2014 22:21:49