COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2016) 1–16

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12396

Words Get in the Way: Linguistic Effects on Talker Discrimination

Chandan R. Narayan,[a] Lorinda Mak,[b] Ellen Bialystok[b]

[a]*Department of Languages, Literatures and Linguistics, York University*
[b]*Department of Psychology, York University*

Received 18 March 2015; received in revised form 6 January 2016; accepted 29 April 2016

Abstract

A speech perception experiment provides evidence that the linguistic relationship between words affects the discrimination of their talkers. Listeners discriminated two talkers' voices with various linguistic relationships between their spoken words. Listeners were asked whether two words were spoken by the same person or not. Word pairs varied with respect to the linguistic relationship between the component words, forming either: phonological rhymes, lexical compounds, reversed compounds, or unrelated pairs. The degree of linguistic relationship between the words affected talker discrimination in a graded fashion, revealing biases listeners have regarding the nature of words and the talkers that speak them. These results indicate that listeners expect a talker's words to be linguistically related, and more generally, indexical processing is affected by linguistic information in a top-down fashion even when listeners are not told to attend to it.

Keywords: Talker discrimination; Indexical processing; Speech perception; Top-down effects; Mental lexicon

1. Words get in the way: Linguistic effects on talker discrimination

The ability to discriminate between individuals solely from their vocalizations is found in many species (e.g., mouse-eared bats, bottlenose dolphins) (Janik, Sayigh, & Wells, 2006; Yovel, Melcon, Franz, Denzinger, & Schnitzler, 2009). In humans, the recognition of individuals from voice alone is determined by a multidimensional suite of acoustic characteristics unique to the talker, including spectral envelope and its change over time, fluctuation in fundamental frequency and amplitude, moments of periodicity and aperiodicity, and long-term averaged spectrum (e.g., Bricker & Pruzansky, 1976; Fant, 1966;

Correspondence should be sent to Chandan R. Narayan, Department of Languages, Literatures and Linguistics, York University, Toronto, ON M3J 1P3, Canada. E-mail: chandann@yorku.ca

Hecker, 1971; Hollien & Klepper, 1984; Xu, Homae, Hashimoto, & Hagiwara, 2013). As it is clear that the speech signal carries both linguistic (e.g., phonological, lexical) information (Allopenna, Magnuson, & Tanenhaus, 1998; Gaskell & Marslen-Wilson, 2002) as well as acoustic correlates to talker identity (e.g., Creel, Aslin, & Tanenhaus, 2008; Nygaard, Sommers, & Pisoni, 1994), a growing psycholinguistic literature seeks to understand the interaction between levels of abstraction of the speech signal and the cognitive processes involved in speech perception, word recognition, and word learning. This research program has produced converging results indicating that low-level acoustic information specific to talkers (or indexical information) affects higher level linguistic processing in a *bottom-up* fashion (e.g., Goldinger, 1996, 1998; Green, Tomiak, & Kuhl, 1997; Mullennix & Pisoni, 1990; Palmeri, Goldinger, & Pisoni, 1993).

Listeners exhibit interference effects from talker information when attending to lexical and phonological aspects of speech stimuli. For example, words are more readily identified when presented by a single talker than multiple talkers (Mullennix, Pisoni, & Martin, 1989), and phonetic perception is adversely affected by increasing talker variability (Green et al., 1997; Mullennix & Pisoni, 1990). These bottom-up effects on linguistic processing are variable, however. Words spoken by the same talker result in a greater recognition memory of words, but this advantage is graded such that the more complex the encoding (phonological or syntactic), the less of a talker effect is observed (Goldinger, 1996). The interaction between talker characteristics and linguistic structure suggests that listeners retain detailed acoustic images of voices and these details affect linguistic representations more than linguistic information affects voice characteristics. When listeners are asked to classify words according to talker gender or initial phoneme, voice variations impair phoneme classification more than the reverse (Mullennix & Pisoni, 1990).

These bottom-up effects observed in word recognition and speech perception highlight the perceptual interruptions that occur when listeners attempt to selectively attend to different aspects of the speech signal without actively encoding talker-specific information in memory. In Vitevitch's (2003) shadowing task experiment in which listeners were asked to repeat spoken words of varying difficulty (in terms of frequency and phonological neighborhood density), at least 40% of participants did not detect a change in talker midway through the task. Listeners' "change deafness" results from their not attending to changed indexical aspects of the stimulus. That is, change-deaf listeners were attending the phonological structures of words rather than lower level indexical information. While some listeners might be better than others at suppressing non-attended information (change-deaf listeners), the suggestion that listeners can completely attend and ignore different aspects of the speech signal is perhaps premature given that the instantiation of linguistic structure is necessarily extracted from acoustic structure that includes indexical features (though see Garrido et al., 2009; for research on *phonagnosics* who understand language but are unable to identify speakers). For example, the results of Remez, Fellowes, and Rubin (1997) suggest that listeners use a single representation to discriminate linguistic and indexical information, whereby attention to phonetic properties (in general) of the speech signal not only aids indexical discrimination but may also aid the recognition of linguistic information.

More recently, the literature on indexical information and its interaction with linguistic knowledge has focused on the ways in which listeners use talker characteristics to aid in the task of lexical access and word learning. Access to semantic information in phonologically dense neighborhoods is affected by talker dynamics. For example, talker variability constrains lexical access such that words that are phonologically similar (e.g., *sheep, sheet*) compete *more* during lexical access when spoken by the same speaker than different speakers (Creel et al., 2008). The acoustics of indexical information might limit the sets of semantic information available to the listener as well (Geiselman & Bellezza, 1976, 1977; Geiselman & Crawley, 1983; Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008). Acoustics can potentially prime semantic information and therefore lead listeners to associate certain forms of semantic knowledge with voices. For example, listeners might have an experience-based expectation that voices with higher or lower fundamental frequencies will express semantic information along gendered distinctions. Creel and Tumlin (2011) refer to these acoustically driven beliefs about the talker's semantic information as *talker-semantic* information. Talker semantics are coupled with *acoustic-matching* effects, whereby listeners recognize more quickly words that had been previously presented and therefore match a stored acoustic image (Goldinger, 1996; McQueen, Cutler, & Norris, 2006; Sjerps & McQueen, 2010).

The emerging pattern of indexical and linguistic processing relationships is one of upstream influences, with talker variability affecting language tasks in ways that promote or hinder linguistic categorization, recall, or learning. The motivation for much of this research has been to understand how the cognitive system extracts linguistic information from a highly variable speech signal. In this paper, we are interested in the related question of whether a basic, low-level cognitive ability such as talker discrimination can be interrupted by the higher level hierarchical structures of language. The interaction between indexical information and linguistic representations is important to explore directly for it allows us to determine the extent to which speech processing is reliant on both the systematic acoustic variation that characterizes different voices and the abstract makeup of words. We address this issue by asking whether listeners can selectively attend to indexical information and ignore linguistic information in a talker discrimination task. Is there an interaction between various types of linguistic information (phonological, lexical, semantic) and talker dynamics in speech processing? To what extent is the processing of talker dynamics affected by the graded linguistic complexity observed in bottom-up processing (Goldinger, 1996; Vitevitch, 2003)? Listeners in our task were asked to discriminate talkers from two serially presented single word utterances that varied along phonological and lexico-semantic dimensions. A previous small-scale study (Babel, McAuliffe, & Narayan, 2012; Narayan, Babel, & McAuliffe, 2014) suggested that, when two words, each spoken by a different speaker, formed a compound word (e.g., "fire"-"man"), listeners' perception of the talker difference was affected. In this study, we pursue a more rigorous test of the interaction between the linguistic relationship between words and the discriminability of their talkers by varying the degrees of semantic relatedness between words (serially presented words that form a compound or a reversed compound) and introducing phonological similarity (rhymes) between words. These four types of word pairs reflect both different levels of linguistic encoding (phonological vs. semantic), as well as different degrees of linguistic relatedness (lexical compound vs. reversed lexical compound vs. words with no lexico-semantic relationship). We hypothesize that these aurally presented words pairs will engage listeners' linguistic systems in a graded fashion, reflecting the various levels and degrees of linguistic abstraction required for understanding speech. The addition of a rhyme condition is motivated by research suggesting that lexical access proceeds in stages to the extent that semantic and phonological levels are engaged differently (cf. Levelt, 1992, for word production; and Pisoni & Luce, 1987, for word recognition). The reversed compound condition is introduced to allow for varying degrees of lexical engagement; that is, as reversed compounds are *less good* lexical compounds, we expect them to have effects different on talker discrimination than do compounds. As words presented serially in pairs are more or less related along these abstract linguistic dimensions of phonology and lexico-semantics, so too do we expect there to be gradient effects on talker discrimination.

2. Method

2.1. Participants

Participants were 116 undergraduate students (51 males, 65 females, M_{age} = 21.9 years) recruited from the undergraduate psychology research pool. Participants were administered the Peabody Picture Vocabulary Test (PPVT) III, Form A. The PPVT is a standardized measure of English receptive vocabulary (Dunn & Dunn, 1997). As the main task in our study assessed interaction between the lexicon and talker discrimination, the PPVT was administered to ensure participants had comparable vocabularies. Participants were asked to indicate which one of four pictures best matched with a verbally presented word. Items were arranged in order of increasing difficulty. Testing ended when the participant made eight errors in a set of 12 items. Standard scores were converted from raw scores based on the participant's age. The standard score has a mean of 100 and a standard deviation of 15. Two participants were removed because of very low receptive vocabulary scores (standard score more than two *SD* below the mean). Five additional participants were removed from the final sample because of very low overall accuracy (overall accuracy more than 2.5 *SD* below the mean) on the main experimental task described below. The final sample consisted of 109 undergraduate students (47 males, 62 females) whose English receptive vocabulary score was within normal range (M_{PPVT} = 96.9, SD_{PPVT} = 10.8).

2.2. Materials

Thirty-two English monosyllabic nouns were chosen representing a range of lexical frequencies from the SubtlexUS corpus (Brysbaert & New, 2009), phonological segments (e.g., consonant manner, voicing of initial segments, etc.), and syllable structures (e.g., CV, CVC, VC). From this base list of target words, four word pair conditions were created: (a) phonological rhymes (e.g., *day-bay*), (b) lexical compounds (e.g., *day-dream*), (c) reversed compounds (e.g., *dream-day*), and (d) unrelated (e.g., *day-bee*). The unrelated pairs were created in a pseudo-random fashion (to avoid rhymes or other linguistic relationships between words) by pairing target words with words from the rhyme list (see Appendix A for the complete list of stimuli). Four native speakers of Canadian-English (2 female, 2 male) recorded the words individually in list form in a sound attenuated recording booth. Speakers were instructed to read the word list, which contained multiple repetitions of stimulus materials ordered randomly, in a monotone list intonation with a 1-second pause between words to avoid compound intonation patterns. Tokens from the end of the list were not used as materials as they were often produced with falling, list-final prosody. Appendix B provides basic voice quality measures for the four speakers based on their productions of the cardinal vowels /i/, /u/, and /a/.

2.3. Procedure

Trials were presented aurally over Sennheiser (HD515) headphones using E-prime presentation software (Psychology Software Tools, Inc.; Pittsburgh, PA, USA). In each trial, the two words in the word pair were separated by a 500 ms ISI. All trials timed out at 2,000 ms at the offset of the second word if no responses were made. The design was fully balanced representing 64 trials per word pair condition: 32 same talker trials and 32 different talker trials. Different talker trials consisted of both same gender and different gender pairs. Participants were instructed to determine, as quickly as possible, whether the two speakers in each trial were the same person or not by pressing keys on opposite sides of a keyboard. Key assignments were counterbalanced. Feedback was only given during the practice trials.

3. Results

3.1. Accuracy

Means for listener accuracy according to Voice Type and Word Type conditions are given in Table 1.

Table 1
Listeners' mean accuracy and standard errors in talker discrimination task according to word type and voice type

	Word Type			
Voice Type	Rhyme	Compound	Reversed Compound	Unrelated
Different	0.75 (0.007)	0.76 (0.007)	0.79 (0.006)	0.80 (0.007)
Same voice	0.87 (0.005)	0.84 (0.006)	0.80 (0.007)	0.73 (0.007)

Listener accuracy was analyzed with mixed-effects logistic regression models (using the *lme4* package in the R statistical environment) (Bates, Mächler, Bolker, & Walker, 2015) with fixed effects of Word Type (Rhyme, Compound, Reversed Compound, and Unrelated) and Voice Type (Same or Different voice) and Subjects and Items as random effects (cf., Jaeger, 2008). There was significant improvement in the model fit when the random effects in the model allowed individual subjects to vary according to the additive Word Type and Voice Type conditions ($\chi^2(14) = 544.23$, $p < .001$). Results of the full model are given in Appendix C.

Figure 1 shows predicted estimates of accuracy based on the mixed-effects model and shows the significant interaction between Word Type and Voice Type ($\chi^2(3) = 260.06$, $p < .001$). Within each Voice Type, pairwise comparisons revealed varying differences in accuracy between Word Types. Table 2 gives difference coefficients between Word Types for each Voice Type condition.

When voices were different, listeners' accuracy to Rhymes and Compounds were similar, and less than their accuracy to Reversed Compounds and Unrelated words. That is, when two words had a linguistic relationship (phonological or lexical), listeners made more errors in determining that two speakers are indeed different.

When voices were the same, listeners' accuracy reflected the graded nature of the relationship between words, with Rhymes resulting in the most accurate responses, followed by Compounds, Reversed Compounds, and finally Unrelated words. Between Voice Types, all similar Word Types were significantly different from each other, except

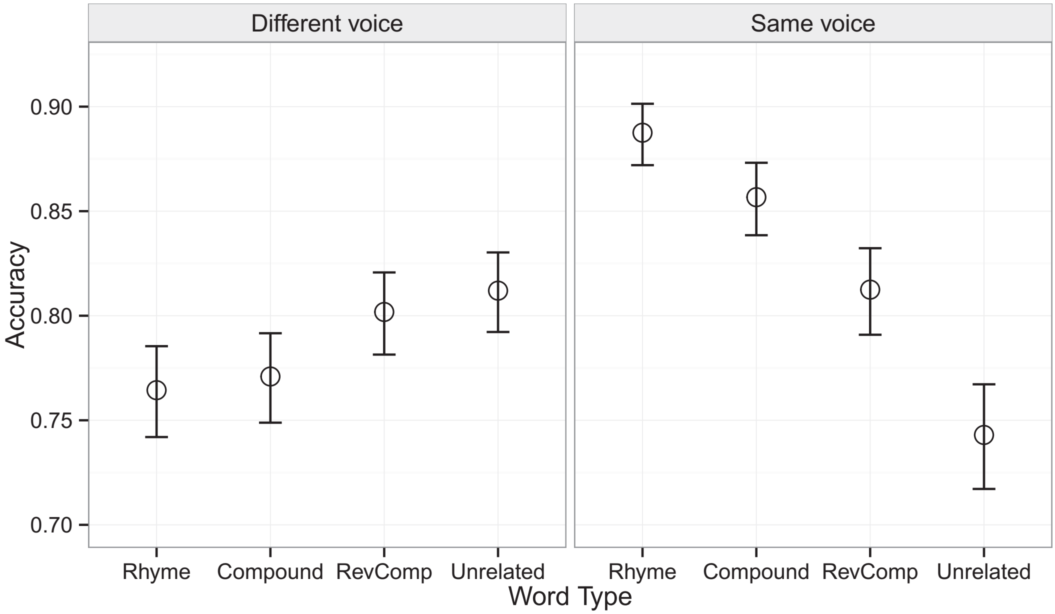

Fig. 1. Predicted estimates of accuracy by Word Type and Voice Type based on the mixed-effects logistic regression model. Error bars represent 95% confidence intervals.

Table 2
Pairwise comparisons of word type and voice type effects on predicted estimates of listener accuracy based on the mixed-effects logistic regression model

Word Type Comparison	Voice Type	β	*SE*	*z*	*p*
Rhyme-unrelated	Different	−0.27	0.07	−4.98	<.0001
	Same	1.00	0.07	14.88	<.0001
Compound-unrelated	Different	−0.23	0.07	−4.31	<.0001
	Same	0.72	0.07	12.18	<.0001
Reversed compound-unrelated	Different	−0.03	0.07	−0.96	.26
	Same	0.39	0.07	7.08	<.0001
Rhyme-compound	Different	−0.02	0.07	−0.59	.53
	Same	0.27	0.08	4.02	<.0001
Compound-reversed compound	Different	0.17	0.07	3.13	<.01
	Same	−0.30	0.07	−4.97	<.0001
Rhyme-reversed compound	Different	−0.20	0.07	−3.87	<.0001
	Same	0.56	0.08	9.12	<.0001

Reversed Compounds, which were not significantly different between same and different voices (β = 0.05, *SE* = 0.09, *z* = 0.63, *ns*).

3.2. *Reaction time*

Table 3 gives listeners' raw reaction times (ms) according to Voice Type and Word Type conditions. Reaction times (RT) to correct responses (n = 23,550) were log transformed for modeling. These data were analyzed using mixed-effects linear regression models (*nlme* package in R) (Pinheiro, Bates, DebRoy, & Sarkar, 2013). The model that specified Subjects (allowed to Word Type and Voice Type) and Items as random effects fit the data significantly better than the more basic model with random intercepts for Subjects (intercept only) and Items ($\chi^2(14) = 225.88$, $p < .001$). Results of the reaction time model are given in Appendix D.

The predicted effects of Word Type and Voice Type on logRT are given in Fig. 2. The figure shows the interaction between the two predictive factors on listeners' reaction times ($\chi^2(3) = 73.22$, $p < .005$).

Table 4 presents difference coefficients for each Word Type and Voice Type resulting from the regression model of logRT.

Table 3
Reaction times (ms) and standard errors for listeners' correct responses according to voice type and word type

	Word Type			
Voice Type	Rhyme	Compound	Reversed Compound	Unrelated
Different	987 (6.95)	966 (6.82)	966 (6.92)	974 (6.67)
Same	998 (5.92)	1,048 (6.57)	1,044 (7.03)	1,088 (7.50)

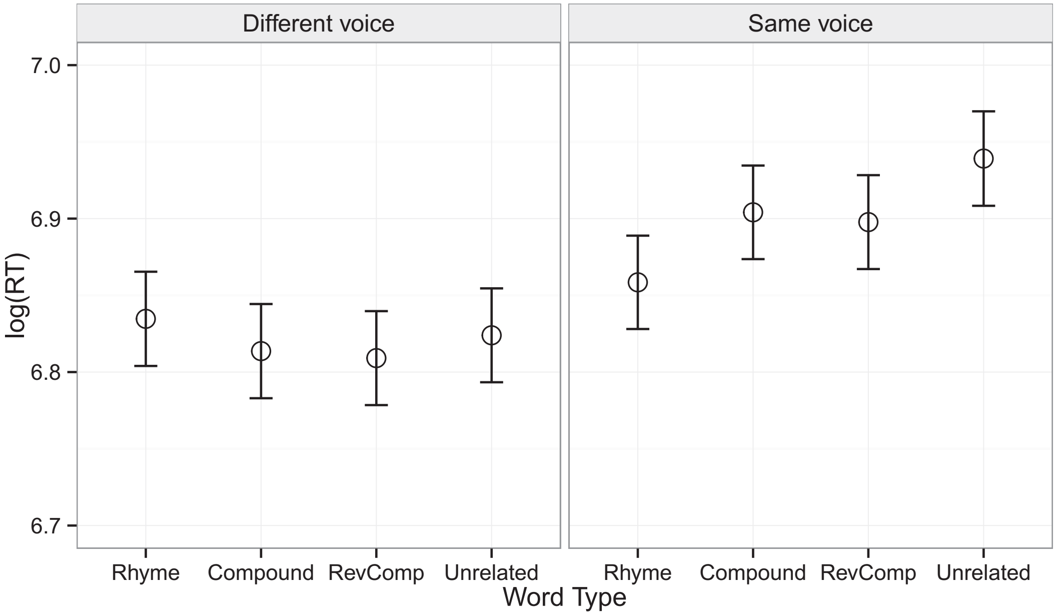

Fig. 2. Predicted estimates of reaction time (log) by Word Type and Voice Type based on the mixed-effects linear regression model. Error bars represent 95% confidence intervals.

Within Voice Types, logRTs showed considerable variability. When talkers were the same, Rhymes were discriminated the fastest, followed by Compounds and Reversed Compounds, which had similar logRTs. Unrelated words had the slowest responses. When talkers were different, logRTs to the four Word Types were not different from one another. Between Voice Types, logRTs for every comparison within a Word Type were significantly different.

Table 4
Pairwise comparisons of word type and voice type effects on predicted estimates of listener reaction time (log) based on the mixed-effects linear regression model

Word Type Comparison	Voice Type	β	*SE*	*z*	*p*
Rhyme-unrelated	Different	0.01	0.01	1.39	.16
	Same	−0.08	0.01	−10.50	<.0001
Compound-unrelated	Different	−0.01	0.01	−1.27	.20
	Same	−0.04	0.01	−4.52	<.0001
Reversed compound-unrelated	Different	−0.01	0.01	−1.89	.12
	Same	−0.04	0.01	−5.28	<.0001
Rhyme-compound	Different	0.01	0.01	1.01	.08
	Same	−0.05	0.01	−6.18	<.0001
Compound-reversed compound	Different	−0.004	0.01	−0.59	.56
	Same	−0.01	0.01	−0.84	.40
Rhyme-reversed compound	Different	0.03	0.01	3.25	.08
	Same	−0.04	0.01	−5.23	<.0001

4. Discussion

The ability to tell two talkers apart solely from serially presented single word utterances relies not only on the salience of the acoustic dimensions characterizing their voices, but also on the linguistic character of their words. Our study shows that listeners, when explicitly instructed to discriminate talkers' voices, unavoidably process the linguistic properties of the talkers' words. Furthermore, these effects on talker discrimination reveal biases that listeners have about talkers and the structures of words they say. Much in the way that listeners might have indexically driven semantic beliefs about the types of words a talker might use, our results show that the converse is also likely. The types of words that talkers use bias listeners' judgment of talkers' voices. Voices that produce a string of words that are linguistically related (either phonologically or semantically) are perceived as indexically similar as well. Thus, the presence of a linguistic relationship between words facilitates the decision that the words were spoken by the same person, whereas the same linguistic relationship interferes with the judgment that the words were in fact spoken by different people. Put another way, relatedness between words along linguistically significant dimensions renders the voices speaking those words more similar.

Facilitation and interference in our task was evident from accuracy measures and, to a lesser degree, reaction time measures, which revealed a graded speech processing according to the degree of linguistic relatedness between words—from strongly related words (rhymes and compounds) to weakly related words (reversed compounds) to unrelated words. These various levels of linguistic information contribute to the overall evaluation process of talker discrimination in a hierarchical fashion, with certain types of linguistic relationships between words being more facilitative of talker discrimination than others. The speed at which listeners responded was less indicative of linguistic effects, but reflected a more general speed-accuracy tradeoff, where listeners were less accurate, but faster, at discriminating different talkers than same talkers.

When talkers were the same, words that rhymed resulted in the highest discrimination accuracy and fastest judgments. This indicates that perceiving two voices as being the same is facilitated by similar phonetic events in the signal, with the most robust confirmation of talker similarity resulting from an acoustic match between the two words. In the rhyme condition, the evaluation is direct and requires the least cognitive effort (evidenced by short RT). The phonological similarity between rhyming words allowed for a comparison between voices, whose acoustic content was as similar as could be without being the same lexical item. This suggests that the acoustic-phonetic structure of the word pairs served as an entrée to listeners' evaluation of talker similarity. To what degree are rhyming words engaging higher levels of linguistic processing in talker discrimination? The abstract or representational interpretation of the rhyme effect is complicated by the acoustic similarity between rhymes spoken by the same talker. That is, listeners need not necessarily abstract a phonological representation of the words to show a facilitative effect of rhymes spoken by the same talker. Consistent with results showing listeners encoding acoustic specifics of speech stimuli as part of a words representation (Creel et al., 2008; McQueen et al., 2006; Sjerps & McQueen, 2010), the rhyme condition in the present task allowed listeners to make a direct comparison of talkers' voices. Phonological similarity allows for acoustic similarity, thus obscuring the interference effect of linguistic abstraction *per se*, but provides evidence for phonologically mediated acoustic discrimination. Listeners' performance with rhyming words is also consistent with Creel et al. (2008), who found that phonologically similar words compete more in a lexical access task when spoken by the same speaker, suggesting a close relationship between talker-specific instantiations of acoustically similar words for the listener.

When talkers were the same, talker discrimination of word pairs that formed compounds was less accurate and slower than when words formed rhymes, suggesting that phonological similarity is more effective than lexical status in communicating talker similarity. Talker identity does receive support from the lexicon, however, as evidenced by the graded accuracy for compounds, reversed compounds, and unrelated word pairs. Listeners more accurately perceived same talkers when words formed compounds than reversed compounds, and reversed compounds than unrelated words. This suggests that listeners have a hierarchically ordered bias of linguistic expectations for the semantic content of words spoken by same talkers. That reversed compounds spoken by the same talker led to discrimination accuracy lower than compounds and higher than unrelated pairs indicates that reversed compounds engage the lexicon. Further support for the linguistic processing of reversed compounds spoken by the same speaker comes from listeners' reaction times, which were no different from the compound condition. The processing of reversed compounds reveals the opportunistic nature of the interface between linguistic structure and indexical dynamics, suggesting that the reversed compound is a plausible enough sequence of words to be spoken by the same talker. Unrelated words spoken by the same talker show the lowest accuracy and slowest reaction time, indicating that listeners' processing of the same voice is disrupted when words have no semantic or phonological relationship.

When talkers were different, overall accuracy was lower than the same voice condition, suggesting that talkers were confusable at an extra-linguistic level. The linguistic effects on talker discrimination accuracy were less graded when talkers were different relative to the same condition. Rhymes and compounds showed similar effects, with accuracy significantly lower than reversed compounds and unrelated pairs. Unlike the same talker condition, when rhyming words are spoken by different talkers, listeners must necessarily represent words phonologically to make a comparison. While rhymes spoken by the same talker are acoustically similar, rhymes spoken by different talkers need not be. This suggests that rhymes spoken by different talkers are engaged linguistically and represented abstractly by the listener. Listeners' perception of rhymes spoken by different talkers is no different from different talker compounds, suggesting a similar linguistically engaged processing whereby phonologically or semantically related words essentially render different talkers' voices more similar. Reversed compounds and unrelated words had a similar effect on discrimination accuracy and reaction time in the different talker condition, indicating that reversed compounds are processed in a way comparable to unrelated words. The semantic relationship between words in reversed compounds changes according to the talker condition—engaging the listener's lexicon only when talkers are the same.

Taken together, our interpretation of the basis for these effects is that listeners' linguistic systems become tuned to the voice of a talker when hearing the first word of the word pair, resulting in a set of phonological and lexico-semantic expectations for the upcoming word. The indexical properties of first word primes listeners' phonology and lexicon for the second word, and this processing is tiered, reflecting the nature of linguistic relationship between the words. The degree of linguistic relatedness between words is reflected in the interference with same-talker accuracy. If the second word in the pair matches the talker-specific phonology (rhymes) or lexicon (compounds and reversed compounds), same-talker discrimination is more accurate than instances where the second word results in unrelated word pair. This talker-specific linguistic processing can render different voices of linguistically related pairs more similar. That is, when linguistic processing is engaged during talker discrimination, indexical acoustic properties are potentially trumped after the presentation of the first word, resulting in a high error rate for linguistically related word pairs spoken by different voices. The practical import of these results is that words compete with voices and can compel the listener to perceive talkers as more or less similar.

5. Conclusion

This study indicates that listeners' phonology and lexicon become tuned to the indexical properties of the talker at the moment of speech processing and, as a result, exert biases on the perception of subsequent speech. Listeners expect two serially presented words to be spoken by the same speaker *and* to be linguistically related. This expectation is much like a converse analog to talker-semantic expectations described by Creel and Tumlin (2011) and others: Listeners' experience with the nature of spoken words informs their understanding of talker identity and linguistic grammaticality (phonological and lexical). Similar to other instances of top-down effects in speech perception (e.g., the "Ganong" effect where real words bias the perception of ambiguous phonetic segments; Ganong, 1980), this study demonstrates the interaction between indexical and linguistic information and the relatively privileged status of words over voices in speech perception. There is no evidence as yet that the same-talker/linguistically related bias is a general feature of all language processing or whether it is a probabilistic outcome of language acquisition and experience. Research with listeners whose lexicons are less mature (e.g., children) than those of listeners in this study will allow us to better understand whether these biases are built-in to language. Furthermore, research into different types of relationships between words (e.g., theta-role assignments in syntactic and semantic relationships) and their talkers will allow us to hone our understanding of the interaction between language and speech.

Acknowledgments

This research was supported in part by a research grant from York University to CN and grant A2559 from the Natural Sciences and Engineering Research Council Canada to EB.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–439. doi:10.1006/jmla.1997.2558

Babel, M., McAuliffe, M., & Narayan, C. (2012). Linguistic effects on talker discrimination: The effect of semantic cohesion. In *Proceedings of LabPhon13*, (p. 235). Stuttgart, Germany.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Bricker, P. D., & Pruzansky, S. (1976). Speaker recognition. In N. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 295–326). New York: Academic Press.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. doi:10.3758/BRM.41.4.977

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*(2), 633–664. doi:10.1016/j.cognition.2007.03.013

Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language*, *65*(3), 264–285. doi:10.1016/j.jml.2011.06.005

Dunn, L. M., & Dunn, D. M. (1997). *Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service.

Fant, G. (1966). A note on vocal tract size factors and non-uniform F-pattern scalings. *Speech Transmission Laboratory Quarterly Progress and Status Report*, *1*, 22–30.

Farrús, M., Hernando, J., & Ejarque, P. (2007). Jitter and shimmer measurements for speaker recognition. Presented at *INTERSPEECH* Antwerp: Belgium.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(1), 110–125. doi:10.1037/0096-1523.6.1.110

Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., & Duchaine, B. (2009). Developmental phonagnosia: A selective deficit of vocal identity recognition. *Neuropsychologia*, *47*(1), 123–131. doi:10.1016/j.neuropsychologia.2008.08.003

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology*, *45*(2), 220–266. doi:10.1016/S0010-0285(02)00003-8

Geiselman, R. E., & Bellezza, F. S. (1976). Long-term memory for speaker's voice and source location. *Memory and Cognition*, *4*(5), 483–489. doi:10.3758/BF03213208

Geiselman, R. E., & Bellezza, F. S. (1977). Incidental retention of speaker's voice. *Memory and Cognition*, *5* (6), 658–665. doi:10.3758/BF03197412

Geiselman, R. E., & Crawley, J. M. (1983). Incidental processing of speaker characteristics: Voice as connotative information. *Journal of Verbal Learning and Verbal Behavior*, *22*(1), 15–23. doi:10.1016/S0022-5371(83)80003-6

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183. doi:10.1037/0278-7393.22.5.1166

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105* (2), 251–279. doi:10.1037/0033-295X.105.2.251

Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception and Psychophysics*, *59*(5), 675–692. doi:10.3758/bf03206015

Hecker, M. H. (1971). Speaker recognition: An interpretive survey of the literature. *ASHA Monographs*, *16*, 1–103.

Hollien, H., & Klepper, B. (1984). The speaker identification problem. In R. E. Rieber (Ed.), *Advances in forensic psychology and psychiatry* (pp. 87–111). Norwood, NJ: Ablex.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. doi:10.1016/j.jml.2007.11.007

Janik, V. M., Sayigh, L. S., & Wells, R. S. (2006). Signature whistle shape conveys identity information to bottlenose dolphins. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *103*(21), 8293–8297. doi:10.1073/pnas.0509918103

Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition*, *42*(1), 1–22. doi:10.1016/0010-0277(92)90038-j

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science*, *30*(6), 1113–1126. doi:10.1207/s15516709cog0000_79

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*, *47*(4), 379–390. doi:10.3758/BF03210878

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, *85*(1), 365–378. doi:10.1121/1.397688

Narayan, C. R., Babel, M., & McAuliffe, M. (2014). Lexical processing masks speaker and writer detail. Paper session presented at the International Conference on the Mental Lexicon, Niagara-on-the-Lake, ON.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42–46. doi:10.1111/j.1467-9280.1994.tb00612.x

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309–328. doi:10.1037//0278-7393.19.2.309

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2013). R Development Core Team (2012). *nlme: linear and nonlinear mixed effects models*. R package version 3.1-103. R Foundation for Statistical Computing, Vienna.

Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, *25*(1), 21–52. doi:10.1016/0010-0277(87)90003-5

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 651–666. doi:10.1037/0096-1523.23.3.651

Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 195–211. doi:10.1037/a0016803

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591. doi:10.1162/jocn.2008.20054

Vitevitch, M. S. (2003). Change deafness: The inability to detect changes between two voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 333–342. doi:10.1037/0096-1523.29.2.333

Xu, M., Homae, F., Hashimoto, R., & Hagiwara, H. (2013). Acoustic cues for the recognition of self-voice and other-voice. *Frontiers in Psychology*, *4*, 735. doi:10.3389/fpsyg.2013.00735

Yovel, Y., Melcon, M. L., Franz, M. O., Denzinger, A., & Schnitzler, H. (2009). The voice of bats: How greater mouse-eared bats recognize individuals based on their echolocation calls. *PLoS Computational Biology*, *5*(6), doi:10.1371/journal.pcbi.1000400

Appendix A: Trial word pairs

Base	Compound	Rhyme	Reversed Compound	Unrelated
Day	Day-dream	Day-bay	Dream-day	Day-bee
Bed	Bed-bug	Bed-bread	Bug-bed	Bed-hook
Wheel	Wheel-chair	Wheel-meal	Chair-wheel	Wheel-math
Light	Light-bulb	Light-kite	Bulb-light	Light-sport
Book	Book-shelf	Book-hook	Shelf-book	Book-house
Paint	Paint-brush	Paint-saint	Brush-paint	Paint-pie
Ice	Ice-cream	Ice-mice	Cream-ice	Ice-key
Sun	Sun-burn	Sun-bun	Burn-sun	Sun-soot
Trash	Trash-can	Trash-rash	Can-trash	Trash-bear
Shoe	Shoe-lace	Shoe-zoo	Lace-shoe	Shoe-fuss
Mouse	Mouse-trap	Mouse-house	Trap-mouse	Mouse-rash
Mail	Mail-box	Mail-snail	Box-mail	Mail-well
Ring	Ring-tone	Ring-king	Tone-ring	Ring-booth
Court	Court-room	Court-sport	Room-court	Court-tack
Bird	Bird-cage	Bird-word	Cage-bird	Bird-core
Air	Air-plane	Air-bear	Plane-air	Air-mall
Hall	Hall-way	Hall-mall	Way-hall	Hall-show
Loop	Loop-hole	Loop-troop	Hole-loop	Loop-bun
Door	Door-knob	Door-core	Knob-door	Door-bay
Cell	Cell-phone	Cell-well	Phone-cell	Cell-saint
Knee	Knee-cap	Knee-key	Cap-knee	Knee-mice
Foot	Foot-ball	Foot-soot	Ball-foot	Foot-bar
Toe	Toe-nail	Toe-show	Nail-toe	Toe-word
Bath	Bath-tub	Bath-math	Tub-bath	Bath-troop
Car	Car-wash	Car-bar	Wash-car	Car-fruit
Boy	Boy-friend	Boy-soy	Friend-boy	Boy-zoo
Tea	Tea-bag	Tea-bee	Bag-tea	Tea-snail
Eye	Eye-brow	Eye-pie	Brow-eye	Eye-kite
Back	Back-ground	Back-tack	Ground-back	Back-soy
Boot	Boot-camp	Boot-fruit	Camp-boot	Boot-king
Bus	Bus-stop	Bus-fuss	Stop-bus	Bus-meal
Tooth	Tooth-paste	Tooth-booth	Paste-tooth	Tooth-bread

Appendix B: Average fundamental frequency (Hz), jitter (%), and shimmer (%) measures for cardinal vowels spoken by two males and two females. Averages were taken from 10 words from the stimulus materials list

	/i/			/u/			/a/		
	F0	Jitter	Shimmer	F0	Jitter	Shimmer	F0	Jitter	Shimmer
Female 1	211	2.6	10.03	229.3	2.18	4.37	209.89	1.78	5.6
Female 2	222	1.1	5.29	264.01	2.32	11.49	226.43	0.88	6.75
Male 1	114	2.05	5.82	169.51	2.43	6.26	121.37	0.85	4.63
Male 2	132	2.42	7.65	123.27	1.76	13.11	132.8	1.58	5.59

Jitter is variation in local fundamental frequency from the speaker's average fundamental frequency. Jitter is perceived as *roughness* in a speaker's voice. Shimmer is variation in cycle-to-cycle amplitude. Shimmer is perceived as vocal *crackle*. Fundamental frequency, jitter, and shimmer measures have been found to characterize voices in speaker verification systems (Farrús, Hernando, & Ejarque, 2007)

Appendix C: Fixed effects and interactions for talker-discrimination accuracy. Reference value for Word Type is Unrelated pairs, and Different voice for Voice Type

	β	*SE*	*z*	*p*
Intercept	1.22	0.07	22.80	<.0001
Word Type$_{\text{Rhyme}}$	−0.25	0.07	−4.85	<.0001
Word Type$_{\text{Compound}}$	−0.24	0.07	−4.24	<.0001
Word Type$_{\text{RevCompound}}$	−0.01	0.07	−0.85	.41
Voice Type	−0.38	0.08	−4.06	<.0001
Word Type$_{\text{Rhyme}}$ × Voice Type	1.28	0.07	14.78	<.0001
Word Type$_{\text{Compound}}$ × Voice Type	0.99	0.07	11.57	<.0001
Word Type$_{\text{RevCompound}}$ × Voice Type	0.44	0.07	5.49	<.0001

Appendix D: Fixed effects and interactions for talker-discrimination reaction times (log). Reference value for Word Type is Unrelated pairs, and Different voice for Voice Type

	β	*SE*	*z*	*p*
Intercept	6.71	0.02	437.44	<.0001
Word Type$_{\text{Rhyme}}$	0.01	0.01	1.22	.20
Word Type$_{\text{Compound}}$	−0.01	0.01	−1.06	.23
Word Type$_{\text{RevCompound}}$	−0.01	0.01	−1.74	.051

(continued)

Appendix D. *(Continued)*

	β	*SE*	*z*	*p*
Voice Type	0.10	0.01	9.89	<.001
Word Type$_{\text{Rhyme}}$ × Voice Type	−0.08	0.01	−5.85	<.005
Word Type$_{\text{Compound}}$ × Voice Type	−0.02	0.01	−2.02	<.05
Word Type$_{\text{RevCompound}}$ × Voice Type	−0.02	0.01	−2.11	<.05