Original Article

Language *and* Speech

Social Expectation Improves Speech Perception in Noise

Language *and* Speech
1–20

DOI: 10.1177/0023830914565191
las.sagepub.com
SAGE

Kevin B McGowan
Stanford University, Stanford CA, USA

Abstract
Listeners' use of social information during speech perception was investigated by measuring transcription accuracy of Chinese-accented speech in noise while listeners were presented with a congruent Chinese face, an incongruent Caucasian face, or an uninformative silhouette. When listeners were presented with a Chinese face they transcribed more accurately than when presented with the Caucasian face. This difference existed both for listeners with a relatively high level of experience and for listeners with a relatively low level of experience with Chinese-accented English. Overall, these results are inconsistent with a model of social speech perception in which listener bias reduces attendance to the acoustic signal. These results are generally consistent with exemplar models of socially indexed speech perception predicting that activation of a social category will raise base activation levels of socially appropriate episodic traces, but the similar performance of more and less experienced listeners suggests the need for a more nuanced view with a role for both detailed experience and listener stereotypes.

Keywords
Speech perception, sociophonetics, exemplar models, variation

Introduction

There is mounting evidence from laboratory phonology and sociophonetics demonstrating that social expectations can influence listener performance on a variety of behavioral measures (Drager, 2010). Manipulating listeners' beliefs about the age, gender, sexual orientation, race, etc. of a talker can lead to behavioral responses which suggest that listeners process the speech signal differently in response to different social expectations. Perception of particular phonetic cues can be altered in response to a primed social group–in other words, listeners appear to dynamically alter the attentional weights associated with particular phonetic cues in response to manipulated social expectations (Pierrehumbert, 2003). This alteration has been shown to occur with cues that reflect both actual usage (e.g., Hay, Warren, & Drager, 2006; Schulman, 1983) and stereotypical usage (e.g.,

Corresponding author:
Kevin B McGowan, Stanford University, Margaret Jacks Hall, Building 460, Stanford CA 94305-2150, USA.
Email: kbmcgowan@stanford.edu

Mack & Munson, 2012). Social perception effects suggest that speech perception proceeds not by winnowing away noise to arrive at a core, intended signal, but by exploiting real patterns of informative variation to impose socially informed structure upon the phonetic signal.

There is evidence in the literature that allophonic and coarticulatory variation can support and enhance perception and word recognition. Sumner (2011) finds that variation is both necessary and beneficial for English listeners overcoming a gross categorical mismatch in French-accented patterns of voice onset time (what Best (1995) classifies as the single category assimilation of two variants). Whalen (1984) demonstrates that listeners are sensitive to subcategorical acoustic mismatches below the level of experimenter awareness. The gating task of Lahiri and Marlsen-Wilson (1991) demonstrates that English-speaking listeners hearing portions of CV and CV stimuli report hearing a CVN, rather than CVC, target when vowel nasalization is present. Beddor, McGowan, Boland, Coetzee, and Brasher (2013) report eye-tracking evidence that coarticulatory cues speed the time course of lexical activation – with English-speaking listeners in a forced-choice visual world task taking immediate advantage of available cues to nasalization. Beddor (2009), citing decades of research on the perceptual consequences of coarticulation, argues that coarticulation provides listeners with informative variation which gives structure to the acoustic signal and simplifies the task of perception. Individual listeners assign different weights to particular acoustic cues consistent with their experience of the variability and usefulness of those cues in understanding speech. Allophonic and coarticulatory variation, then, have been shown to both support and enhance perception and word recognition. The position taken in this paper is that patterns of *socially meaningful* variation should prove to be similarly useful for listeners.

If listeners use patterns of socially informative variation available in the speech stimulus, and if this invokes a set of specific phonetic expectations that covary with contextual cues, then manipulating social expectations should be able to *enhance*, not merely alter, perception of an acoustic stimulus. Szakay, Babel, and King (2012) used a cross-language and cross-dialect priming task to demonstrate that variable sociophonetic cues can facilitate priming between translation equivalents. Sumner and Samuel (2009) showed that listeners with experience with a particular accent–Long Island English–enjoyed greater semantic priming by that accent than listeners who reported lacking this experience. In contrast, Sumner and Kataoka (2013) found that General American listeners, even those who see no benefit from the non-rhotic Long Island-accented primes, *do* see a benefit from similarly non-rhotic Southern British English primes–suggesting that simple exposure to a phonetic cue is not the sole determining factor in how richly experience with that cue is encoded and stored in memory, but that encoding of phonetic experience is socially weighted (Sumner, Kim, King, & McGowan, 2014). In this paper, I will present evidence from an experiment in which listeners were asked to transcribe Chinese-accented speech in noise. The results of this experiment suggest that manipulation of social expectations about a given talker can enhance listeners' ability to disambiguate accented speech in noise. Listeners are more accurate when transcribing Chinese-accented speech in noise when shown an ostensible speaker with a Chinese face than when presented with a Caucasian face. This performance is compared to results in a control condition in which the guise is a stylized human silhouette.

The present result provides support for the usefulness of patterns of socially meaningful variation, and therefore provides support for a model of speech perception in which listeners' linguistic representations encode not only this variation, but also link this variation to social categories. This finding is inconsistent with phonological models and theories of speech perception in which listeners are presumed to abstract away from detailed knowledge of the signal to an idealized or underspecified form. This outcome provides support for models in which listeners can link social category information to particular acoustic cues (Sumner et al., 2014). While exemplar models certainly have both of these affordances (Johnson, 1997, 2006; Munson, 2010), it does not

necessarily follow that listeners store detailed episodic traces of experience and draw upon these experiences during perception. One can imagine, for example, extending the double-weak model (Nearey, 1992, 1997) to include a set of social cues whose settings alter the weights assigned to particular acoustic cues to segmental and prosodic identity. Indeed, these affordances are consistent with any model of speech perception in which it is conceded that listeners are capable of learning. This could, in principle, include storage-intensive K-nearest-neighbor type algorithms, to which exemplar models are closely related, or simple Bayesian models which presume the storage of no training data at all (Norris & McQueen, 2008; Staum Casasanto, 2009).

One means of testing the particular claims of exemplar models in a behavioral task is to control for listener experience in such a way that one group of listeners has demonstrated detailed knowledge of a particular speech variety and a second group has demonstrated a comparative lack of this knowledge. In the experiment described here, this is implemented by dividing listeners into either a "more experienced" group of listeners who have a high level of experience with Chinese-accented English and a "less experienced" group with a low level of experience with this variety. The result of this manipulation is surprising in that experienced and less experienced listeners appear to benefit approximately equally from the presence of accurate social information about the speaker. This issue is explored further in the Discussion and Conclusion sections.

1.1 Ideology or experience?

An alternative to this view that listeners use available cues to talker social category to simplify the task of perception (by invoking either attentional cue weights or patterns of phonetic experience consistent with that social category) is the bias hypothesis offered by Rubin (1992). The bias hypothesis, further developed in Lippi-Green (1997) and named "reverse linguistic stereotyping" in Kang and Rubin (2009), holds that the results of social speech perception experiments may be best understood as emerging from listeners' language ideologies and biases. The matched guise task presented here is, essentially, a replication of the first experiment in Rubin (1992). Rubin's study played identical mini-lectures to undergraduate students who were shown a photograph of either an Asian or a Caucasian graduate student instructor who they were led to believe was the speaker on the audio tape. Rubin found that listeners who saw the Asian photograph perceived the voice to be more strongly accented. These listeners also tended to have lower scores on a cloze test–although this difference was not statistically significant.

Listeners who believed the instructor to be Asian tended to retain particular lexical items less well. Rubin interprets this finding as evidence of negative social bias on behalf of the listeners –specifically, due to a lack of homophyly. These negative attitudes toward Asian instructors lead, Rubin argues, to a communicative breakdown in the classroom separate from any legitimate claims about difficulty understanding a non-native speaker's accented English. Regardless of how hard the non-native instructor may have worked to achieve a native-like English accent (indeed, even if this effort were completely successful), students' firm belief that foreign instructors will be incomprehensible will still result in perceptions of hallucinated accentedness and thus reduced perceptibility.

This interpretation was later endorsed and extended by Lippi-Green (1997). Lippi-Green introduces the concept of "communicative burden." Listeners and speakers are engaged in the shared act of communicating. Speakers control, to the best of their ability, how much energy they expend producing speech that is maximally clear to the listener, but listeners are not merely passive receivers. Listeners control how much energy *they* are willing to expend in decoding the signal, resolving ambiguous segments, etc. Lippi-Green endorses Rubin's interpretation that listeners will perceive an accent even when one is not present, but adds to this the notion that the

listener is ultimately culpable for the resulting communication breakdown. In this view, negative social bias leads listeners to choose to expend less energy decoding the acoustic signal, resolving ambiguous segments, etc.

When shown an Asian face, listeners in Rubin's study are, according to Lippi-Green, shirking their portion of the communicative burden. Following Rubin's own interpretation, Lippi-Green claims Rubin's findings indicate that "preconceptions and fear are strong enough motivators to cause students to construct imaginary accents, and fictional communicative breakdowns." (Lippi-Green, 1997, p. 128). The assumption underlying these claims is that Rubin's results indicate reduced attention on the part of listeners in the Asian instructor condition. Due to racial bias, these listeners are simply not attending to the acoustic signal as closely as those in the Caucasian face condition.

Such an interpretation stands in stark contrast to the model in which listeners are actively using social knowledge (including social stereotype) to structure and interpret the acoustic signal. It may be possible to reconcile these positions by noting that Rubin's participants fared more poorly in the incongruous or mismatched condition (Asian instructor paired with Standard American English voice). It may be the case, although it is difficult to discern from Rubin's results, that listeners in the congruous condition (American instructor paired with Standard American English voice) either responded at ceiling or enjoyed enhanced perception due to the presence of the supporting socially meaningful prime. Listeners in the incongruous condition, led to believe the graduate instructor was from China, may have been primed to listen with inaccurate expectations of Chinese-accented English. One can imagine this mismatch between social expectation and acoustic signal resulting in degraded performance.

To test for precisely this effect, the current study employs, alongside congruous and incongruous conditions, a control condition in which listeners are presented with neither accurate nor inaccurate social information about the speaker. To ensure that listeners in this guise are still motivated to think about the identity of the speaker but with no cues to that identity, and to keep the visual and auditory structure of the trials as similar as possible across guise conditions, this control condition was implemented using a simple, geometric silhouette.

1.2 Bias or incongruence?

As noted above, the present experiment is intended to be essentially a replication of Rubin (1992) to address both lingering questions about the interpretation of that result and the advances in theories of sociophonetic perception. In Rubin's experiment, participants heard recordings of Standard American English and images were used to manipulate their beliefs about the racial identity of the speaker. In the experiment reported here, listeners hear recordings of Chinese-accented English and different faces are displayed to shift social expectations during transcription. However, the alignment of congruous and incongruous face and voice pairs has been inverted from Rubin's design. Specifically, in Rubin (1992), those seeing an Asian face reported expecting an accent that was not, in fact, present in the audio recordings. Those seeing a Caucasian face did not expect a foreign accent and did not hear one. In the present study, we might hypothesize that those seeing an Asian face expect a non-native accent. The voice in the recorded sentences does, indeed, have a Chinese accent. Listeners seeing a Caucasian face will anticipate a Standard American English accent but will hear, instead, Chinese-accented speech.

If listeners in the present experiment who believe the speaker to be Chinese transcribe recordings more accurately than those who believe the speaker of the same recordings to be Caucasian then it seems fairly clear that, contrary to the bias/reverse linguistic stereotyping hypothesis, expectation of even a stigmatized foreign accent can have a facilitatory effect on the understanding

of accented speech. Implications for the usefulness of social knowledge during speech perception will be the central question of the subsequent discussion. Also, as discussed above, a silhouette condition, intended to convey no social information, is included to help distinguish between facilitation when the face and voice support one another and inhibition when there is a face/voice mismatch–a control missing from the Rubin study. Listeners are grouped by reported experience level with authentic Chinese-accented English to test the hypothesis that social effects in perception provide strong evidence in favor of exemplar theories of speech perception. Because level of experience with a particular variety of accented English may be unavailable to conscious introspection, a separate attempt was made to assess the accuracy of these self-reported groupings.

2 Methodology

Eighty-seven listeners from the University of Michigan and the University of California, Berkeley participated in a transcription in noise task. Each listener was presented with an image of their purported talker that remained on screen throughout the experiment. They heard a series of 60 English sentences read by a native speaker of Mandarin Chinese and embedded in noise. Listeners were asked to type, as accurately as possible, what they heard.

2.1 Stimuli

Stimulus materials consisted of 30 pairs of high and low predictability sentences originally developed by Bradlow and Alexander (2007). Bradlow and Alexander created the high predictability sentences using an iterative sentence completion paradigm with groups of non-native and native speakers of English. Sentences in the high predictability list are those that consistently received the most consistent completion results from both populations. The low predictability sentences replace the semantically informative material with uninformative frames. The pairing of high and low predictability sentences should allow us to gauge any contribution of social knowledge to sentence perceptibility over and above the better-understood contribution of semantic knowledge. The Wildcat Corpus (Van Engen et al., 2010) contains high quality recordings of these sentences by a number of native Mandarin speakers. The recordings used in this experiment were read by a 23 year old female Chinese native speaker of Mandarin (Wildcat Corpus speaker CHF02). The full set of sentences used is listed in appendix A.

The scripted recordings from the Wildcat Corpus were segmented into individual sentence-length files and equated in amplitude. These files were then mixed with native English multitalker babble (Van Engen & Bradlow, 2007) using the sox audio processing tool to create speech-in-noise recordings with a – 4 dB signal-to-noise ratio at the target word. This signal-to-noise ratio was determined after a series of pilots using the full set of sentences with no noise in which participants across conditions demonstrated a clear ceiling effect in transcription accuracy. An informal listening task completed by several researchers unfamiliar with the semantic content of the sentences suggested that mixing 76 dB noise with a 72 dB signal made the transcription task sufficiently difficult to compare error rates.

Multitalker babble was selected over white, Brownian, or other possible types of noise to enhance the ecological plausibility of the stimuli for participants. Listeners in this task are being asked to draw on their social expectations under laboratory conditions; these more random types of noise created stimuli that seemed, to the experimenter, to be more clinical and less natural-sounding.

The target words themselves occur uniformly in sentence-final position with the falling intonation typical of English declaratives and with the declination typical of the end of a prosodic group.

Figure 1. Faces used in the transcription experiment.

This speaker was chosen from the set of available speakers, in part, because there is no obvious list intonation in her reading of the scripted sentences. Beyond this uniformity, the target items represent a varied set of vowels, consonants, consonant clusters, number of syllables, and morphological complexity.

The actual target norms for L2 English learners in China have traditionally been British rather than American English (Kirkpatrick & Zhichang, 2002), though there may be a shift currently underway to American English norms in textbooks and pedagogical recordings. This fact surely influences the English acquired by Chinese learners and may interact with and shape American listener expectations about Chinese-accented English. The belief that a speaker of Chinese English will be non-rhotic, for example, may well be attributable to this legacy.

Prior to the presentation of the experimental stimuli, listeners heard and transcribed four practice items intended to capitalize on recognizable associations between face, accent/voice, and semantic content. The goal of these practice items was both to make participants comfortable with the transcription user interface and, implicitly, to reinforce that face and voice might be meaningfully linked in the experiment. Listeners saw an image of the face of the speaker and transcribed, in random order, two recordings of Leonard Nimoy as the character Spock and two recordings of Arnold Schwarzenegger speaking characteristic lines of dialogue. Practice items were also presented in multitalker babble.

2.2 Visual stimuli

The transcription task described here is an inverted matched guise task. Matched guise is a well-established experimental technique in sociolinguistics for teasing apart auditory indexical information (Campbell-Kibler, 2005; Lambert, Hodgson, Gardner, & Fillenbaum, 1960) and perceived social properties. The present experiment is "inverted" because it presents visual stimuli to establish social expectations and then measures the extent to which these social expectations can influence the perception of phonetic detail and word recognition in noise.

One of three images was presented to listeners to establish these social expectations; the faces are shown in Figure 1. Each listener saw only one of the three images (between-subjects design) and the image was displayed for the duration of each trial. The Asian and Caucasian images were found via a web search for license-free portraits and, beyond an informal survey, have not been normed for attractiveness, racial typicality, gender stereotypicality, memorability, etc. It is quite

likely that differences along these dimensions exist–the Caucasian image shows someone who is more clearly smiling, for example–but it is unclear why or how such differences might impact listeners' perception of Chinese-accented speech either positively or negatively.[1] The silhouette condition is extremely minimalistic. This is the result of iterative development from what was originally a blue silhouette created by tracing photographs. In a series of pilot experiments, participants were able to identify the race of these outline silhouettes at levels well above chance. To avoid any confounds due to racial perceptions of the silhouette, the current radically simplified representation was chosen instead. It is worth noting that this simplification also removes other, potentially useful, social cues such as gender, age, and size of the speaker. However, any effect this loss of information might have on perception should affect comparison with the two face conditions equally.

2.3 Participants

Eighty-seven undergraduate students participated at one of two experiment sites: the University of Michigan phonetics lab or the University of California, Berkeley phonology lab.

2.3.1 Michigan listeners. Fifty-seven undergraduate students from the University of Michigan Introductory Psychology subject pool participated for partial course credit. Participants had no known hearing problems. The initial intention was to recruit groups of listeners with both high and low levels of experience with Chinese-accented English from this population. However, only five participants with extensive experience could be identified. These five participants have been excluded from analysis and a second experiment site at the University of California, Berkeley was added (see below). In all, seven Michigan participants were excluded prior to data analysis–five for reporting extensive experience with Mandarin Chinese, either through language study or, in four cases, by being bilingual. One participant was excluded for browsing the web and sending text messages on his smartphone during the experimental session. One final participant was excluded from the data analysis for struggling to remain awake during the experiment and reporting the task as extremely difficult. Race and ethnicity information for these participants was not collected. Data for 50 participants are reported here: 16 in the Asian face condition, 16 in the Caucasian face condition, and 18 in the silhouette condition.

2.3.2 Berkeley listeners. As indicated in the previous section, locating a sufficiently large population of Chinese-English listeners at the original University of Michigan research site proved problematic. The goal was to identify a population of listeners who could be expected to have not only copious experience hearing Chinese-accented English, but also accurate linkages between this experience and a social category of 'Chinese' that might be activated by seeing a face. To achieve this, it was determined early in the design of the experiment that the target participants for the more experienced group should be heritage Mandarin speakers who report growing up in a home with Mandarin speaking parents or grandparents but who themselves speak little or no Mandarin. This situation, one might presume, should lead to extensive experience with Chinese-accented English.

One might also imagine recruiting as experienced listeners students who have taken one or more courses from an L1 Mandarin speaking instructor. On one hand, this is an appealing population as they are relatively easy to locate at the University of Michigan and may have accurate knowledge that their instructor was truly from China. On the other hand, Rubin and Lippi-Green hypothesize a confound for our purposes with native English speakers exposed to Chinese-accented English through L1 Chinese professors and graduate student instructors. If these monolingual English

listeners truly are refusing to attend to their Chinese-accented instructors then they may fail to encode their voices and thus, despite exposure, not represent an experienced population.

Ultimately, 31 heritage Mandarin undergraduate students from the University of California, Berkeley participated in exchange for US$15.00 each. Three sets of data were not analyzed: one L1 speaker of Mandarin had misunderstood the flier, a second participant misrepresented his identity, and a third data file was overwritten prior to analysis due to experimenter error. All participants in this group were Chinese Americans. In all, 10 participants were randomly assigned to the Asian face condition, eight to the silhouette condition, and 10 to the Caucasian face condition.

2.3.3 Assessing experience. For all participants, experience level was assessed in two ways: through self-report on an 11 question survey and through a yes/no accent authenticity detection task described in detail in McGowan (in press). Along with questions about birthplace, places lived, languages spoken at home, and languages spoken personally, the survey asked participants to agree or disagree, on a five point Likert scale to: having experience listening to Chinese-accented English; to having close friends who speak Chinese as a first language; to having family members who speak Chinese as a first language; and a number of questions intended to ascertain listener ideologies (e.g., "It is socially acceptable to imitate a Chinese accent" or "I can distinguish a Chinese accent from a Korean or Japanese accent"). Listeners selected as "less experienced" (i.e., the Michigan listeners) had a mean age of 19, had lived 98% of their lives in the United States, were predominantly born in Michigan (51%) or New York (18%), reported having no friends or family members who spoke Chinese as a first language, and on average claimed not to have a clear idea what a Chinese accent sounds like or to be able to distinguish Korean or Japanese-accented English from Chinese-accented English. Listeners selected as "more experienced" (i.e., the Berkeley listeners) had a mean age of 22, had lived 82.3% of their lives in the United States, were predominantly born in California (55%) or China (38%), reported having friends and/or family members who spoke Chinese as a first language, reported Mandarin as a language spoken at home, and on average claimed to have a clear idea what a "Chinese" accent sounds like and to be able to distinguish Korean or Japanese-accented English from Chinese-accented English. Both groups tended to agree that it is socially unacceptable to imitate a Chinese accent, with one participant writing in, "Unless you're Asian!"

In addition to this assessment of participants' conscious awareness of Chinese and Chinese-accented English experience, listeners participated in an accent authenticity detection task intended to gauge their ability to discriminate authentic Chinese-accented English voices from the voices of American English speakers producing imitated Chinese accents. The explicit assumption motivating this accent authenticity detection task is that one's ability to accurately discern authentic from imitated Chinese-accented English improves with increased exposure to the authentic variety. Following Johnson (2006), this should be particularly true when the listener is aware that the speaker is Chinese and is therefore able to form a correct link between the social category and the stored experience. Listeners with less experience, by contrast, prefer imitations of a target accent to authentic examples. Indeed, it is entirely possible that imitated varieties of accented speech make up much of the less experienced listeners' exposure to what they conceive of as the target variety. Neuhauser and Simpson (2007) found that German monolingual speakers were more likely to identify German imitations of French and American accents as authentic than they were to correctly identify true non-native accents.

Neuhauser and Simpson (2007) argue that listeners and accent imitators share a surprisingly uniform cognitive prototype of what features an imitated variety should have, while true non-native speakers produce patterns which defy these expectations. Therefore, not only should listeners with experience listening to the authentic variety be better able to make use of non-stereotypical

Table 1. More and less experienced listener authenticity ratings for authentic and imitated Chinese-accented English.

Accent	Experience	Proportion rated "authentic"	SE
Authentic	More	0.64	0.009
	Less	0.36	0.007
Imitated	More	0.065	0.005
	Less	0.19	0.006

markers of authenticity and language specific patterns of coarticulation in the authentic variety (e.g., Beddor, Harnsberger, & Lindemann, 2002), but listeners with less experience should be more drawn to the imitated accent. As shown in Table 1, the self-reported more experienced listeners were significantly better than the self-reported less experienced listeners at discriminating authentic from imitated Chinese-accented English ($\beta = .03978; t = 2.594; p < 0.05$). In other words, more experienced listeners were less likely to be misled by the imitation. Accuracy scores were calculated for each participant in this accent ratings task and will be considered in the statistical analysis of the transcription experiment. For a fuller exploration of this task, which participants completed immediately after the transcription in noise task described in the remainder of this paper, please see McGowan (in press). The stimuli described in the remainder of this paper were recorded by a native speaker of Mandarin Chinese reading English sentences; no further use will be made of imitated materials.

2.4 Procedure

Participants at the University of Michigan research site used Apple Macbook Computers (model 4,1; late 2008) in a sound-attenuated booth at the University of Michigan, Department of Linguistics; stimuli were presented over AKG K271 mkII headphones.

Participants at the University of California, Berkeley research site used the same laptop computers in a quiet space dedicated to speech perception experiments in the Cal Phonology Lab. This was not a sound-attenuated booth. AKG K240 headphones were used in place of the AKG K271 mkII.

While every effort was made to maintain consistency across these two experimental locations, there were a number of factors that could not be controlled (e.g., Michigan listeners were in a sound booth, Berkeley listeners were in a quiet room; Michigan listeners received course credit, Berkeley listeners received cash, etc.). In general, these differences do not seem to provide either group with an unfairly easier task, but the possibility does exist.

Prior to their arrival, participants were randomly assigned to one of the three guise conditions: Asian face, silhouette, or Caucasian face. Responses were entered via the Macbook keyboard and listeners were instructed to advance trials using the return key to minimize trackpad use. Stimuli were presented using Superlab stimulus presentation software version 4.0.8. Volume was set at a comfortable listening level. Listeners were told they would listen to the voice of a graduate student instructor and that they would see the face of this talker throughout the experiment. The talker would either be a native speaker of Mandarin Chinese or a native speaker of American English. It was further explained that photographs were not available for all graduate student instructors so some participants, the experimenter was not sure who, would see a placeholder silhouette image instead.

The instructions provided for the transcription portion of the task informed participants that they would hear 60 sentences in "cocktail party noise" and that their task was to type, as carefully

Table 2. Predicted tendencies for each model in visual conditions.

Condition	Exemplar model	Bias model
Silhouette	baseline	baseline
Asian (congruous)	enhancement	inhibition
Caucasian (incongruous)	inhibition	baseline

as possible, what they heard. It was explained that they would hear each sentence only once and could not repeat it, but to listen closely and type whatever words they were able to make out. It was further explained that they would hear four practice sentences to get comfortable using the program and that, while spelling counted, speed did not so they should take their time.

2.5 Measurements and statistical analysis

Participants' transcriptions of the authentically Chinese-accented speech in noise were automatically normalized to lowercase, stripped of any punctuation and automatically coded as correct or incorrect using a Python script. This script set a Boolean "isCorrect" variable to true if the target word came at the end of the participant's response and false otherwise. These automated decisions were reviewed by a research assistant who was naive to the goals and design of the experiment. A small number of coding decisions were reversed for being mere typographical errors (e.g., "coffe" for *coffee* or "yello" for *yellow*). An analysis of the most frequent transcription errors will be presented in the "Results" section below. A response was coded as correct only if it contained the target (final) word in the sentence; a response satisfying this criterion could be otherwise blank or contain gibberish and still be coded as "correct."

These coded transcription responses were then analyzed using the open source statistical package R (R Development Core Team, 2011). The data were modeled with linear mixed-effects models, as implemented in the **lme4** (Bates, Maechler, & Bolker, 2014) R package. Categorical variables were sum-coded to allow the interpretation of lower order effects in the models as main effects rather than simple effects. Models were fitted with the maximal random effects structure justified by model comparison and the data to avoid the inflated risk of type 1 errors in random intercept-only models (Barr, Levy, Scheepers, & Tily, 2013). Model comparison was also used to justify the inclusion of fixed effects and interaction terms in the linear models while statistical significance within the resulting models will be reported using Satterwate's approximations as implemented in the R package **lmerTest** (Kuznetsova, Brockhoff, & Christensen, 2014).

2.6 Predictions

In the Introduction, two models that make different predictions about the results of this task were proposed for listeners' use of social primes. These predictions are summarized in Table 2. The first model is an exemplar model like that described in Johnson (2006), while the second is the bias model initially proposed by Rubin (1992). According to the predictions of a socially informed exemplar model, being provided with social cues about the talker should raise the base activation level of stored episodic traces consistent with the social categories indexed by those cues. This raised activation has the effect of causing less frequently experienced phonetic sequences to be perceived as more categorical than they might normally be (Johnson, 2006, p. 494). The silhouette provides no cues to talker identity and so should not raise base activation levels, resulting in baseline transcription accuracy for the listener. The congruous condition, when a Chinese face cues

social categories consistent with the Chinese-accented voice, should raise base activation levels of any stored exemplars consistent with those social categories and thus result in enhancement of transcription accuracy. The incongruous condition, when a Caucasian face cues social categories inconsistent with the Chinese-accented voice, should raise base level activations of episodic traces consistent with a Caucasian social category and therefore result in inhibition–causing the Chinese-accented phonetic information to be perceived less categorically than it might normally be.

The bias model, on the other hand, holds that Caucasian American English-speaking listeners reject their portion of the communicative burden when confronted with cues indicating an Asian non-native speaker. The silhouette again provides no cues to talker identity and so should induce no negative bias, providing baseline transcription accuracy for the listener. The congruous condition displays an Asian face and so should induce negative bias on the part of the listener, resulting in a reduction of attention and encoding resulting in reduced transcription accuracy. Finally, the incongruous condition, which displays a Caucasian face, should induce no negative bias, resulting in transcription accuracy equivalent to the baseline control condition.

2.7 Levels of experience

The two levels of experience are included to probe more deeply the predictions of an exemplar model. Listeners with high and low levels of experience with a particular variety should have correspondingly frequent or infrequent episodic traces of speech that are consistent with that variety. These differences in frequency should lead to stronger, more robust expectations for more experienced listeners and weaker, more specific expectations for less experienced listeners in precisely the way that high and low lexical frequency predict historical lenition patterns in Pierrehumbert (2001).

Exemplar models therefore predict differences between more and less experienced listener populations on this task. The higher frequency of experiences for the more experienced listeners should lead to more detailed and therefore more robust expectations and should result in a stronger enhancement of transcription accuracy. The less experienced listeners, with less frequent experiences of Chinese-accented speech, should suffer from a kind of over-fitting by which a sparse, less robust exemplar cloud is activated by social expectations.

2.8 Levels of predictability

The two levels of predictability are included, in part, to allow comparisons between the magnitudes of any enhancement due to social priming and the better-understood effects of semantic priming on the transcription of speech in noise.

Additionally, the levels of predictability are included to further tease apart the exemplar and bias models. While one can easily imagine an exemplar model that would include a role for context in raising base activations of semantically related words, the bias model offers no such affordance. Listeners in the bias model are said to reject their portion of the communicative burden. The strong interpretation of this hypothesis would predict no benefit for semantic predictability on transcription of a final target word. Even in the absence of this worst case scenario, though, one should expect to see evidence of reduced attendance to the acoustic stimuli causing a reduction in transcription accuracy for biased listeners.

Finally, given the usefulness of semantic information in predicting words in noise, social priming should have little benefit to offer in predictable sentences. Social facilitation should therefore be strongest in the low predictability sentences where the reactivation of appropriate phonetic expectations provided by socially meaningful cues can provide the most benefit.

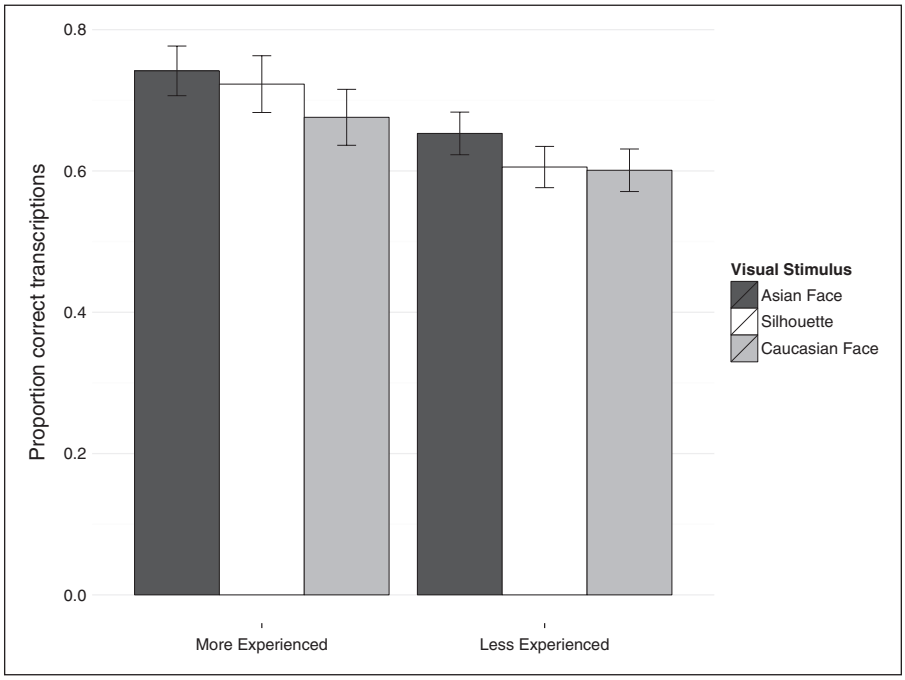

Figure 2. Proportion of correct transcriptions by more and less experienced listeners under the three face conditions.

3 Results

Figure 2 shows the proportion of correct responses for all listeners in each face condition by experience level. A generalized linear mixed model with a logit linking function with a categorical correct/incorrect dependent variable was fitted to the data. Face, predictability, and the self-reported experience levels are included as fixed effects; subject and target word are included as random effects with both random intercepts and random slopes in terms of two of the four predictor variables: experience and predictability, the inclusion of these random slopes having been justified by means of model selection and the data.

As described in the Introduction, the rating score variable was determined via a yes/no accent authenticity detection task completed by each participant immediately after participation in the transcription task. This score encodes each subject's false positives (inaccurately labeling the imitated variety as authentic) subtracted from their hits (accurately detecting the authentic variety). This score is essentially a non-normalized version of a signal detection *d*' score and is used here because six of the more experienced listeners performed at ceiling on the authenticity detection task and produced no false positives, resulting in undefined *d*' values. A simplified model containing only rating score as a fixed effect was compared to a model containing only self-reported experience levels. The model with rating score did not provide a significantly better fit for the data than the model with self-reported experience levels. Since these predictors are highly correlated (0.75) and, at least in theory, assess the same property of each participant (experience with Chinese-accented English), only self-reported experience was included in the remaining models as an assessment of that experience.

Table 3. Proportion of correct transcriptions grouped by semantic predictability in each face condition.

	High predictability	Low predictability
Asian face	0.79	0.58
Silhouette	0.73	0.55
Caucasian face	0.72	0.53

As this figure suggests, there is a significant main effect of face – with transcription in the Asian condition being significantly more accurate than transcription in the Caucasian condition ($\beta = -0.38; t = -2.4; p = 0.0177$). From the perspective of this reference level of "Asian" in the face condition, transcription accuracy appears to be numerically better than in the silhouette condition, but this trend is not significant ($\beta = -0.24; t = -1.4; p = 0.1653$). Judging visually from Figure 2, the silhouette condition, which was intended to provide a baseline, appears to pattern differently for the more experienced and less experienced listener groups. For the more experienced listeners, the silhouette patterns with the Asian face performance, while patterning with Caucasian face performance for the less experienced listeners. However, while there is a significant main effect of experience – with more experienced listeners being overall more accurate transcribers of Chinese-accented English across face conditions ($\beta = -0.82; t = -3.69; p < .001$) – there is no interaction of face and experience. When a full model with an interaction term of face and experience is compared with a reduced model with no interaction term, the addition of the interaction is not significant. There was also no justification for including an interaction term for experience × predictability as it added nothing to the model, $\chi 2(2) = 0.2455; p = .88$.

Table 3 presents transcription accuracy by semantic predictability and face condition. There is a significant main effect of predictability ($\beta = -1.64; t = -3.7; p < .001$), with target words in high predictability sentences transcribed, on average, 19.4% more accurately than targets in low predictability sentences. Transcribers, regardless of experience level, transcribe the final target words more accurately when the preceding sentences contain semantically related words than when the preceding sentence does not provide cues to the final word. As such, the predicted interaction of face and predictability is not observed, $\chi 2(9) = 1.56; p = 0.46$.

3.1 Transcription errors

It seems likely that specific transcription errors, to the extent that these represent mishearings, could reveal influence of social expectations on perception. Table 4 presents the top 10 misheard words sorted from most often to least often misheard. Mishearings are presented for each misheard word divided by face condition and experience level of the listeners who typed the errors. The most frequently misheard word by far, representing nearly 3% of the total mishearings with 131 errors, was "sport," which both listener experience groups transcribed almost universally and regardless of face condition as "spot." For the other nine words, the table lists the top two or, in the case of ties, three mishearings for the target word. For most targets, the transcription errors follow a Zipf-like distribution with a few common error types (those presented in Table 4) and a long tail of singletons.

4 Discussion

One implication for the usefulness of social knowledge during speech perception is clear. The results on this task are clearly inconsistent with the bias model of Rubin (1992) and Lippi-Green (1997). It

Table 4. Top 10 misheard words and their most common mishearings by face condition and listener experience level.

	Asian face		Silhouette		Caucasian face	
	More exp.	Less exp.	More exp.	Less exp.	More exp.	Less exp.
sport	spot	spot	spot	spot	spot	spot
sleeves	leaves	leaves	plate	leaves	leaves	leaves
	sleep	lips	leaves	names	slaves	lips
cents	steps	punched	test	steps	yourself	fence
	sets	fence	steps	fence	steps	steps
bomb	phone	phone	phone	bone(s)	phone	phone
	bunk	bone	bond	ball	uniforms	taught
coach	college	couch	couch	college	college	couch
	couch	college	college	couch	couch	college
grass	graph	graph	graph	graph	ground	graph
	bread	ground	grasps	ground	graphs	ground
father	mother	brother	brother	brother	farther	mother
	bus	wrist	mother	helped	bus	brother
necks	next	next	next	accent	next	Xanax
	nest	x	nest	Xanax	excessively	next
story	stars	star	stories	stars	stories	stars
	stories	dolly	starry	diary	history	diary
sheets	street	trees	trees	trees	trees	trees
	sheep	street	shit	things	sheep	things

is not the case that presenting a Chinese face to listeners results in a reduction in transcription accuracy. There is no evidence of a rejection of the communicative burden, even on the part of the less experienced listeners. There is some evidence, however, that listeners perform less well when presented with an incongruent social prime/voice combination. The most straightforward interpretation of the Rubin (1992) finding, then, is that listeners in that task were presented with misleading social cues that indicated a Chinese-accented voice when a Standard American English voice was presented. Under that interpretation, the lower transcription accuracy for the Caucasian face would appear to replicate the lower cloze test scores for listeners in Rubin's Asian face condition. At the same time, listeners in Rubin's study had surely experienced native speakers of varieties of American English with Asian facial features so it remains an open question whether it was something about the task itself that caused this pairing to behave as a mismatch or whether this behavior represents a common expectation, or stereotype, on the part of Caucasian Americans.

However, Rubin's is not the only evidence presented in the literature for a bias hypothesis. In one classic study, Niedzielski (1999) played recordings of a native Detroit, Michigan speaker for two groups of Detroit listeners. There is a widely held ideology among Michigan speakers that their variety of English is "unaccented" or identical to Standard American English (Niedzielski, 1995, 1997). Though both groups of Detroiters heard identical recordings of Detroit-accented speech, one group was led, in an incongruous condition, to believe the speaker was Canadian while the second group was led, in a congruous condition, to believe the speaker was a fellow Detroiter. When asked to match what they had heard to tokens with resynthesized vowels, the listeners who believed the speaker to be Canadian perceived more Canadian raising than listeners who–correctly–believed the speaker to be from Detroit. Listeners who believed the speaker

to be a fellow Detroiter were also less likely–though not significantly–to choose resynthesized vowels consistent with Detroit speakers' general participation in the northern cities chain shift (Milroy & Gordon, 2003). Instead, these listeners, in both the Canadian and Detroiter conditions, showed a preference for resynthesized vowels unlike the northern cities tokens they had heard and closer to a model of Standard American English. While Canadians do not, in fact, participate in the northern cities shift, Detroiters most certainly do (Labov, 1994). Therefore, these Detroit listeners consistently made incorrect vowel identification decisions in both the congruent and incongruent conditions. Niedzielski (1999) interprets this result as evidence that language ideologies can interfere with listeners' access to fine phonetic detail during perception. One possible interpretation of Niedzielski's study in light of the present result is that listeners do use social expectation to enhance and give structure to the phonetic signal, but that simplifying and structuring the acoustic signal does not necessarily always mean providing greater access to fine phonetic detail.

In line with this interpretation of Niedzielski, Schulman (1983) presented Swedish listeners who were highly proficient in American English with synthesized sVt continua. When instructions were provided in Swedish, listeners could not distinguish the *set*/*sat* portions of the continuum. When instructions were provided in English, however, this contrast was reliably perceived. Listeners' contextually motivated expectations suggested alternate listening strategies. These are strategies which, in real world listening situations outside the speech perception laboratory, may greatly simplify the task of perception by reducing the set of phonetic features that must be attended to in real time to those which have a high probability of being informative. Niedzielski's listeners, from this perspective, are not attending to fine phonetic details beyond those required to ascertain category membership because they have not previously experienced a need to. These features of Detroit speech had not, as Niedzielski points out in her conclusion, risen to the level of conscious awareness among Detroiters and so may not have particularly drawn the attention of listeners.

The implications of the results presented here for socially indexed exemplar models of perception seem somewhat more complicated. In the most general sense, these models are a better fit for the results than the bias model. Just as laid out in the Predictions section for these exemplar models, listeners transcribe Chinese-accented speech more accurately when the face they are shown provides social information congruent with the voice they are listening to. They transcribe speech less accurately when the social cues and acoustic signal are incongruent. The simplest interpretation of this result is that a congruent socially informative prime facilitates transcription and is therefore consistent with the predictions of exemplar models.

However, human beings are not simply passive recorders of data from the peripheral sense organs. Even in the simplest exemplar model, episodic traces will not be stored unless the listener minimally attends to the sensation and encodes it. This point is of particular importance when discussing social expectation because the linkages between social category and episodic trace must depend not on some objective fact about the social categories of the speaker, but on the listener's beliefs about and awareness of those social categories. As such, there should be a difference in the accuracy of the social category "Chinese" for the more experienced listeners in this experiment than for the less experienced listeners. The less experienced listeners surely have some experience with Chinese-accented English, real or imitated, but whether this has been linked in their memories to a specific Chinese social category seems quite unlikely. The more experienced listeners should be able to activate Chinese-accent appropriate exemplar clouds more accurately, with a more specific linkage between stored episodic traces and a Chinese social category. Less experienced listeners, on the other hand, are likely to activate not only their experiences with Chinese-accented English but also with Japanese-accented English,

with Korean-accented English, and with any other accent that the listener's ideologies construe as "Asian."

It should be surprising to exemplar theorists that there is no significant difference in the degree of improvement for more experienced and less experienced listener populations. There is no interaction between face and the self-reported category labels of more and less experienced. The authenticity detection scores, while correctly predicting group membership, are also not significant predictors of transcription accuracy. There should be, a strong interpretation of a Johnson-like (2006) model suggests, frequency effects of both exposure to this variety and awareness of the social category of that exposure for the more experienced listeners. Such an effect is not reflected in the results.

Furthermore, it is not at all clear that *facilitation* is the most accurate description of the differences between the guise conditions observed in these results. Both more and less experienced listeners show higher transcription accuracy in the Asian face condition, but they may not do it by the same mechanism. The tendency for the silhouette control condition to pattern with the Caucasian face for less experienced listeners and with the Asian face for more experienced listeners suggests that the listeners' default expectations about speaker identity are active during presentation of the silhouette. In the absence of informative social cues, less experienced listeners behave as if they anticipate a voice that might be consistent with a Caucasian face while more experienced listeners behave as if they anticipate not only a voice that might be consistent with a Chinese face, but indeed a Chinese-accented voice linked to a Chinese face. One caution about this interpretation that must be given, though, is that the more experienced listeners in this task knew that they had been recruited because they were Chinese Americans with heritage Mandarin experience. This could easily have influenced their default assumptions about the identity of the speaker in the silhouette condition. Bearing that caveat in mind, it does appear that the manipulation of social information tends toward enhancement in the Asian guise for the less experienced listeners condition and toward inhibition in the Caucasian guise for experienced listeners.

Evidence from patterns in the transcription errors (Table 4) is mixed regarding the extent to which listeners made use of social knowledge to alter their specific predictions about fine phonetic detail. While it is true that transcription is overall more accurate in the Asian face condition, it would be better support for this specific hypothesis if transcription errors reflected a kind of compensation for accented articulation. In both high and low predictability contexts, for example, listeners frequently heard *sport* as *spot* (the most common mishearing overall). Sport, in the stimulus recordings, was produced without a clear post-vocalic consonantal [ɹ]. The absence of post-vocalic [ɹ] is a strongly stereotypical feature of Chinese-accented English. If listeners were using either their experience with Chinese or their stereotypes of Chinese to anticipate accented speech we would expect them to anticipate this missing consonant. This is especially true given that the speaker produces a fairly rhotacized vowel in this token. On the other hand, one might also expect to see high front lax unrounded ([ɪ]) percepts in *sleeves* or *sheets* and the mishearings *lips* (*sleeves*) and *shit* (*sheets*) do occur. Perhaps a future replication of this task with stimuli specifically chosen to target a confusability matrix of common L1 Mandarin difficulties with L2 English would be more informative on this point.

Finally, the fact that the self-reported experience levels provided an equally good model fit for the transcription accuracy results tells us that, at least for assessing listeners' experience with Chinese-accented English, self-reporting provides useful insight into listeners' experience. The accent authenticity detection task provided no more insight into listeners' experience than simply asking them. This equivalence may not hold for other varieties of accented English (e.g., AAVE, New York City English), and the accent authenticity detection task may yet prove useful for experience questions that are less available for introspection.

5 Conclusion

The bias hypothesis offered by Rubin (1992) and Lippi-Green (1997) is not consistent with the results presented here. It is not the case that monolingual English speakers presented with a purportedly Asian-faced speaker tune out that speaker or refuse to uphold their end of the communicative burden. In the present experiment, just as in the Rubin study, when the face provided the listener with informative information about the identity of the speaker, performance was improved. Conversely, when the displayed face provided the listener with misleading information about the identity of the speaker, again *just* as in Rubin, performance was lower. Given the improved performance of less experienced listeners on this task, it seems likely that listener stereotypes of Chinese-accented English play a role in speech perception, but that role is demonstrably not a negative one. None of this is intended to suggest that there is *not* negative bias against non-native English speakers in general and perhaps particularly against graduate student instructors. The claim here is simply that this bias does not appear to be implemented at the level of attention or speech perception.

The results reported here are inconsistent with a theory of speech perception that presumes abstraction away from variation in the signal and toward an idealized representation. The presence of a contextual cue to the social identity of the speaker enhances perception of a congruously accented voice in noise. In this respect, the general predictions of exemplar models are clearly upheld. Other evidence in support of exemplar models of perception comes from the fact that experienced listeners are overall better at the task than less experienced listeners. In a sense, this difference replicates the usual finding in the sociophonetic perception literature, though to a lesser degree than might be expected. Furthermore, the patterning of the silhouette condition with the experience level of the listener, while not significant, seems to offer another hint that speech perception in the absence of clear social cues relies heavily on the detailed experience of the listener.

One difficult question only partially addressed in this study but ultimately inescapable in this type of sociophonetic perception work is the role played by awareness, on the part of each individual listener, of the linkages between the detailed phonetic cues to social information and the social categories themselves. To put this plainly, the listener may have detailed experience with a particular variety but attribute it to the wrong social category–or attribute it to a social category that is ultimately either too broad or too narrow (Munson, 2010). For example, an inexperienced listener may have a single broad category of "Asian" into which detailed experience of such disparate varieties as Mandarin-accented English, Cantonese-accented English, Japanese-accented English, and Korean are all combined. Furthermore, the detailed experience itself may be of imitated or stereotypical performances of these various varieties. Ultimately, these distinctions do not matter to current exemplar models. From an exemplar perspective, all that matters is that the listener has a social category that can be activated by some contextual cue and a linkage from that category to a cloud of episodic traces. From that perspective, the present result is consistent with existing exemplar models. However, it is curious that social exemplar dynamics do not appear to demonstrate, at least in this one study and on this one task, the kind of gradient sensitivity to frequency that can be observed with, for example, lexical frequency.

Whether a socially informative prime enhances or interferes with listeners' use of phonetic detail is a function of acoustic cue, social category, listener identity, and context. Exemplar accounts can model these relationships, but results so far do not motivate the storage of episodic traces to the exclusion of other possible explanations. A naive Bayes classifier, with no storage requirements beyond the current state of expectation probabilities, can also model this relationship. Nearey's (1997) double-weak model, extended to include relationships to social variables, could similarly model the present results. Ultimately, the findings here suggest that detailed laboratory characterization of

listeners' experience is desirable to adequately evaluate and advance an exemplar theory of speech perception.

Acknowledgements

This paper has benefited from insightful discussions with Pam Beddor, Steve Abney, Ben Munson, Robin Queen, Julie Boland, Lauren Squires, Susan Lin, Anna Babel, my students at Rice University, and audiences at the Linguistics Society of America and the Acoustical Society of America. I would like to express my sincere thanks to all of the participants who took part in this study and to Keith Johnson for hosting the UC Berkeley sessions. Finally, I am grateful to reviewers Katie Drager and Laura Staum Casasanto, and to Associate Editor Paul Warren, for their patience and valuable suggestions.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Note

1. However, much research in social psychology describes an "attractiveness halo effect" by which physical appearance significantly influences social perceptions and interactions (e.g., Zebrowitz & Franklin, 2014).

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Maechler, M., & Bolker, B. (2014). *Lme4: Linear mixed-effects models using S4 classes*. Retrieved from http://CRAN.R-project.org/package=lme4

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*(4), 785–821.

Beddor, P. S., Harnsberger, J., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, *30*(4), 591–627.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, *133*(4), 2350–2366.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange, & J. J. Jenkins (Eds), *Cross-language speech perception* (pp. 171–204). Timonium, MD: York Press.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, *121*(4), 2339–2349.

Campbell-Kibler, K. (2005). Listener perceptions of sociolinguistic variables: The case of (ING). PhD thesis, Palo Alto, CA: Stanford University Department of Linguistics.

Drager, K. (2010). Sociophonetic variation in speech perception. *Language and Linguistics Compass*, *4*(7), 473–480.

Hay, J., Warren, P., & Drager, K. (2006). factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, *34*, 458–484.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J. W. Mullennix (Eds), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, *34*, 485–499.

Kang, O., & Rubin, D. L. (2009). Reverse linguistic stereotyping: Measuring the effect of listener expectations on speech evaluation. *Journal of Language and Social Psychology*, *28*, 441–456.

Kirkpatrick, A., & Zhichang, X. (2002). Chinese pragmatic norms and 'China English'. *World Englishes*, *21*(2), 269–279.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. (2013). Lmertest: Tests for random and fixed effects for linear mixed effect models (R package version 2.0-6). Retrieved from http://CRAN.R-project.org/package=lmerTest

Labov, W. (1994). *Principles of linguistic change: Internal factors*. Oxford, UK; Cambridge, MA: Wiley-Blackwell.

Lahiri, A., & Marlsen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *3*, 245–294.

Lambert, W. E., Hodgson, R., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *Journal of Abnormal and Social Psychology*, *3*, 44–51.

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. London, UK: Routledge.

Mack, S., & Munson, B. (2012). The influence of /s/ quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp' stereotype. *Journal of Phonetics*, *40*(1), 198–212.

McGowan, K. B. (in press). Sounding Chinese and listening Chinese: Awareness and knowledge in the laboratory. In A. M. Babel (Ed.), *Awareness and control in sociolinguistics*. Cambridge, UK: Cambridge University Press.

Milroy, L., & Gordon, M. (2003). *Sociolinguistics: Method and interpretation*. Oxford, UK: Wiley-Blackwell.

Munson, B. (2010). Levels of phonological abstraction and knowledge of socially motivated speech-sound variation: A review, a proposal, and a commentary on the papers by Clopper, Pierrehumbert, and Tamati; Drager; Foulkes; Mack; and Smith, Hall, and Munson. *Journal of Laboratory Phonology*, *1*, 157–177.

Nearey, T. M. (1992). Context effects in a double-weak theory of speech perception. *Language and Speech*, *35*(1–2), 153–171.

Nearey, T. M. (1997). Speech perception as pattern recognition. *The Journal of the Acoustical Society of America*, *101*(6), 3241–3254.

Neuhauser, S., & Simpson, A. P. (2007). Imitated or authentic? Listeners' judgements of foreign accents. In *ICPhS XVI* (pp. 1805–1808). Saarbrücken, Germany: International Congress of Phonetic Sciences.

Niedzielski, N. (1995). Acoustic analysis and language attitudes in Detroit. In M. Meyerhoff (Ed.), *(N) Waves and means: University of Pennsylvania working papers in linguistics* (Vol. 3, pp. 73–86). Philadelphia, PA: University of Pennsylvania Press.

Niedzielski, N. (1997). The effect of social information on the phonetic perception of sociolinguistic variables. PhD thesis, University of California, Santa Barbara.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*(1), 62–85.

Norris, D., & McQueen, J. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds), *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: Benjamins.

Pierrehumbert, J. B. (2002). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds), *Probability theory in linguistics* (pp. 177–228). Cambridge, MA: MIT Press.

R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rubin, D. L. (1992). Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education*, *33*(4), 511–531.

Schulman, R. (1983). Vowel categorization by the bilingual listener. *PERILUS Working Papers* III: 81–99.

Staum Casasanto, L. (2009). Experimental investigations of sociolinguistic knowledge. PhD thesis, Stanford University Department of Linguistics.

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, *119*, 131–136.

Sumner, M., & Kataoka, R. (2013). Effects of phonetically-cued talker variation on semantic encoding. *The Journal of the Acoustical Society of America*, *134*(6), EL485–EL491.

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

Sumner, M., & Samuel, A. G. (2009). The role of experience in the processing of cross-dialectal variation. *Journal of Memory and Language*, *60*, 487–501.

Szakay, A., Babel, M., & King, J. (2012). Sociophonetic markers facilitate translation priming: Maori English GOAT – A different kind of animal. *University of Pennsylvania Working Papers in Linguistics*, *18*(2), 137–146.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native- and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53*(4), 510–540.

Van Engen, K., & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *Journal of the Acoustical Society of America*, *121*(1), 519–526.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception and Psychophysics*, *35*, 49–64.

Zebrowitz, L. A., & Franklin, R. G. (2014). The attractiveness halo effect and the babyface stereotype in older and younger adults: Similarities, own-age accentuation, and older adult positivity effects. *Experimental Aging Research*, *40*(3), 375–393.

Appendix A. High and low predictability sentences.

High predictability	Low predictability
Elephants are big **animals**.	He pointed at the **animals**.
A pigeon is a kind of **bird**.	We pointed at the **bird**.
The war plane dropped a **bomb**.	Dad talked about the **bomb**.
A quarter is worth twenty-five **cents**.	He pointed at the **cents**.
We heard the ticking of the **clock**.	She looked at the **clock**.
The team was trained by their **coach**.	We read about the **coach**.
Many people like to start the day with a cup of **coffee**.	Mom pointed at the **coffee**.
February has twenty-eight **days**.	There are many **days**.
Last night, they had beef for **dinner**.	He talked about the **dinner**.
My parents, sister and I are a **family**.	We read about the **family**.
A race car can go very **fast**.	She thinks that it is **fast**.
The good boy is helping his mother and **father**.	Mom pointed at his **father**.
People wear shoes on their **feet**.	Mom looked at her **feet**.
When sheep graze in a field, they eat **grass**.	Dad pointed at the **grass**.
I wear my hat on my **head**.	She pointed at her **head**.
At breakfast he drank some orange **juice**.	Mom looked at the **juice**.
In spring, the plants are full of green **leaves**.	She talked about the **leaves**.
People wear scarves around their **necks**.	She talked about their **necks**.
For dessert, he had apple **pie**.	Mom talked about the **pie**.
She made the bed with clean **sheets**.	Dad talked about the **sheets**.
Rain falls from clouds in the **sky**.	Dad read about the **sky**.
The sport shirt has short **sleeves**.	He looked at the **sleeves**.
Football is a dangerous **sport**.	This is her favorite **sport**.
A book tells a **story**.	We looked at the **story**.
A wristwatch is used to tell the **time**.	This is her favorite **time**.
Birds build their nests in **trees**.	He read about the **trees**.
He washed his hands with soap and **water**.	We talked about the **water**.
Monday is the first day of the **week**.	This is her favorite **week**.
Bob wore a watch on his **wrist**.	He looked at her **wrist**.
The color of a lemon is **yellow**.	Mom thinks that it is **yellow**.