



***Thank you for using our service!***

**Interlibrary Services  
The Ohio State University Libraries  
(614)292-6211  
liblend@osu.edu**

Article Express documents are delivered 24/7 directly to your ILLiad account from scanning libraries around the world. If there is a problem with a PDF you receive, please contact our office so we might report it to the scanning location for resolution.

NOTICE

**WARNING CONCERNING COPYRIGHT RESTRICTIONS**

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

## 2 Sounding Chinese and Listening Chinese: Awareness and Knowledge in the Laboratory

---

*Kevin B. McGowan*

### **Perception, Variation, and Awareness**

Perception is the use of knowledge to make sense of – to impose structure upon – the signals relayed to the brain from our peripheral sense organs. For speech perception, this has been construed as using our knowledge of language to impose linguistic structure upon primarily auditory, but also visual (McGurk and MacDonald 1976) and haptic (Gick and Derrick 2009), sensations. The study of speech perception, then, can be framed as the quest to understand the process by which listeners perform this mapping from sensory events to subjective experiences of those events. An important component of this investigation is understanding the knowledge, and the sensory cues to that knowledge, that listeners bring to bear on shaping the sensory events of speech. A theoretical assumption made early in the history of this research was the separation of the speech stream into linguistic<sup>1</sup> cues, useful in lexical access and retrieval, and social cues, useful in obtaining information about the speaker.

Much work on speech perception and word recognition has assumed this separation and taken the linguistic cues, whether acoustic or articulatory, to be both the primary object of research and the primary source of information for listeners. But speech is highly variable, both within and across talkers, so a long-standing puzzle for speech perception researchers has been how listeners could possibly map such highly variable acoustic input onto such apparently consistent subjective experiences of those inputs. This is known explicitly as the “lack of invariance” problem because there appear to exist no sensory cues that map one-to-one to listeners’ subjective experience of particular linguistic units. Lisker (1986), for example, identifies no fewer than sixteen cues sufficient to invoke a subjective experience of voicing in English; none of which is necessary for that percept. Even when not stated explicitly, the notion that invariance really *should* exist at some level commonly emerges

<sup>1</sup> “Linguistic” is a loaded term as it embraces the assumption that social cues are *not* linguistic, but I will use it throughout this chapter as it remains the standard term for referring to sub-phonemic, phonemic, and lexical forms.

in descriptions of variation as a daunting challenge that listeners must somehow overcome in perception.

However, the dual linguistic and social aspects of speech are inescapably conveyed by means of a single acoustic signal. The same phonetic cues that have been long documented as carrying distinctive linguistic meaning simultaneously carry the distinctive social meanings that link a voice to speaker attributes ranging from age to race to sexual orientation (e.g. Foulkes and Docherty 2006; Eckert 2008; Johnstone and Kiesling 2008; Campbell-Kibler 2009; Munson 2010). It seems likely, therefore, that listeners would use not only linguistic but also social knowledge to impose structure upon the sensory events of speech. Perhaps the probability of mapping a particular auditory event to a particular subjective experience of that event is conditional upon the social categories that are contextually salient (Sumner *et al.* 2014).

Indeed, facilitated by the introduction of exemplar models of perception into linguistics (Johnson 1997; Drager 2010, Drager and Kirtley, this volume), sociophonetics and speech perception researchers have consistently found that perception of particular phonetic cues can be altered in response to primed social categories. Listeners appear to dynamically alter the attentional weights associated with particular phonetic cues in response to manipulated social expectations (Pierrehumbert 2002). This alteration has been shown to occur both with cues that reflect actual usage (e.g. Schulman 1983; Niedzielski 1999; Hay *et al.* 2006) and with stereotypical usage (e.g. Mack and Munson 2012, Carmichael, this volume). Socially cued perception effects suggest that speech perception proceeds not by winnowing away noise to arrive at a core, intended signal, but by exploiting listeners' knowledge of real patterns of informative variation to impose structure upon auditory sensations. These findings place speech perception researchers in the unexpected position of needing to quantify and understand not only listeners' linguistic knowledge, but also their social knowledge.

It is typically assumed, following the model offered by Johnson (2006), that listeners whose percepts can be altered by social information necessarily have knowledge both of the socially meaningful phonetic variation itself and of the covarying social category information. In this work, it has seemed methodologically intractable to assess and quantify this knowledge so the argument becomes circular – listeners whose behavior is consistent with exemplar predictions of detailed experience are assumed to have detailed exemplars; listeners whose behavior is not consistent with exemplar predictions are assumed not to have the necessary experience. Escaping this circularity requires us to develop a better understanding of what it means for a listener to be more experienced, how to assess and quantify that experience, and how the accuracy and detail of one's social expectations might shape encoding and memories of linguistic experience.

We know that listeners can and do make use of patterns of socially informative variation in the speech stimulus. Furthermore, manipulating socially cued expectations can enhance, not merely alter, perception of an acoustic stimulus. Dahan *et al.* (2008) show that listeners use their knowledge of [æ]-raising before voiced velar stops, and not before voiceless velar stops, to more quickly distinguish, for example, *back* from *bag* when *bag* in the stimulus set is produced by a speaker with the raised vowel variant. Szakay *et al.* (2012) used a cross-language and cross-dialect priming task to demonstrate that variable sociophonetic cues can facilitate translation priming. Beddor *et al.* (2013) found that listeners have knowledge of informative patterns of nasal co-articulation in American English and can use these patterns to make lexical decisions as soon as evidence of nasalization is provided by the speech stream. These patterns of co-articulation are language-specific (Beddor *et al.* 2002), but also highly individual. Nevertheless, they are useful and informative to listeners. Sumner and Samuel (2009) showed that listener experience with Long Island English predicts the usefulness of semantic priming by voices with that accent. Speakers of Long-Island-accented English saw semantic priming benefits from primes spoken in this variety, while general-American listeners did not. However, Sumner and Kataoka (2013) found that general-American listeners, even those who see no benefit from non-rhotic Long-Island-accented primes, *do* see a benefit from similarly non-rhotic Southern British English primes – suggesting that simple exposure<sup>2</sup> to a variety is not the sole determining factor in how richly experience with that variety is encoded and stored in the lexicon.

That listeners use socially informative variation raises the important question, not normally explicitly addressed in the speech perception literature, of listener awareness. We generally assume in experimental work that speech perception occurs beneath the level of conscious awareness and is unavailable to introspection. Listeners cannot discuss or control whether they perceive consonants categorically (Lieberman *et al.* 1957), perceive non-words as existing words (Ganong 1980), make use of co-articulatory evidence (Mann and Repp 1981), perceive non-native contrasts in terms of native ones (Best 1994), or any number of other classic effects. These linguistic effects occur below the level of awareness and control, so we investigate listeners' knowledge of them by designing other tasks that allow us to measure the influence of these effects indirectly. We assess the influence of semantic relatedness, for example, by measuring how long it takes listeners to decide if a target word is a real word or a non-word when the prime is semantically related or unrelated to that target.

<sup>2</sup> "Simple exposure" here appears to be identical to *perceiving* in the sense of Drager and Kirtley (this volume).

Listeners have knowledge of linguistic effects without necessarily – indeed, without typically – having awareness of that knowledge. This observation has a clear parallel in Squires’s discussion (this volume) of *perceiving* and *noticing* of social variation. Although *perceiving* is probably not the ideal word, Squires has made the novel observation that there is a distinction to be made between perception that occurs when the perceiver is aware of the relationship between sensory cues and social category and perception that occurs when the perceiver is unaware of this relationship. There appear to be at least four distinct pools of knowledge language users are able to use – each with its own gradient awareness. In perception we have knowledge of phonetically cued linguistic information, but also phonetically cued social information. Similarly, for speech production we have knowledge of linguistic information conveyed by speaking, but also knowledge of social meanings conveyed by speaking. Speakers’ gradient awareness of each of these four areas of knowledge has so far received very little attention in the speech perception literature.

Social knowledge has primarily received attention in sociolinguistics. For production, Labov (1972) has described the cline of social awareness as ranging from indicators – features occurring below the level of consciousness – to markers – above the level of consciousness and used to convey and infer social meaning – to stereotypes – describing features above the level of consciousness which are available for meta-linguistic, often stigmatizing, commentary. Stereotypes are strongly linked to speakers’ conceptualization of a variety, but may not actually occur in that variety.<sup>3</sup> Preston (1996, this volume) captures this disjunction and describes clines of awareness for both production and perception by defining awareness along four independent continua: availability, accuracy, detail, and control. *Availability* is concerned with how likely a member of a speech community is to discuss a feature; features may be unavailable – never discussed – or highly available – common topics of conversation. *Accuracy* describes how closely community members’ language ideologies align with linguistic reality; Michigan speakers, for example, commonly believe that their daily speech is without shibboleth and similar to, if not the definition of, supra-local standard American English (Niedzielski 1999). *Detail* describes the specificity of description community members have available to them. Detail ranges from global language ideologies without access to linguistic detail (e.g. “British people sound intelligent”) to specific details (e.g. “Chinese speakers often leave out the definite article” or “Japanese speakers have trouble with r and l”). Finally, *control*, which is explicitly about production, describes whether community members can use

<sup>3</sup> These stereotypes may not occur either because they are inaccurate, such as the expectation of a gay lisp (Mack and Munson 2012), or because a change from above has led to avoidance of the stereotype, as predicted by Labov.

or perform a particular feature (although see Babel (this volume) for a model in which avoidance of a form, non-production, may also signal control).

Conceptualizing listener awareness in this way, as listeners' gradient ability to notice covarying patterns between their linguistic and social knowledges of speech, allows us to develop ways to probe and assess listeners' socially cued knowledge of speech independent of self-reported experience levels. If we suspect that listeners might bring both social and linguistic information to bear on the task of perception, or even if we simply aim to take seriously the predictions of usage-based models of speech perception (Docherty and Foulkes 2014), it becomes necessary either to control listener social experiences for the purposes of laboratory research or to quantify those experiences with tasks that transcend listener awareness of social variation. Controlling listeners' social experience in the laboratory could, in principle, be done by extending artificial language learning. Researchers would invent a language, invent a social context for that language, and train participants with varying levels of experience with those constructs. Methodologically, this route is likely to be extremely time consuming. This approach also has the shortcoming of being infelicitous for the investigation of existing sociolinguistic variables. Finally, it seems likely that participants will understand the invented language and social context in terms of their pre-existing experience with their own native language(s) and culture(s), so, even with an artificial language learning paradigm, the facts of listener experience would still need to be assessed. The alternative approach, quantifying listener experience, is still a daunting logistical challenge, but has the benefit of allowing investigation of existing sociolinguistic patterns using listeners' life experiences.

The goal of the present chapter is to explore the feasibility of assessing and quantifying listeners' accumulated linguistic and social experiences for modeling experience in laboratory research. Experiment 1 presents an accent authenticity detection task designed to indirectly measure listeners' knowledge of Chinese-accented English in terms of their ability to discriminate authentic Chinese-accented English voices from the voices of American-English speakers producing imitated Chinese accents. This task is intended to measure listener knowledge of phonetically cued social information without depending on listeners having conscious awareness of that knowledge. The explicit assumption is that one's ability to accurately discern authentic from imitated Chinese-accented English improves with increased exposure to the authentic variety. Listeners accumulate experience with covarying patterns of phonetic and social information. More experienced listeners should have a greater number of detailed experiences with patterns of phonetic variation, even those, such as Labov's indicators, below the level of conscious awareness or unavailable for introspection or commentary. Less experienced listeners, on the other hand, should have fewer detailed experiences with these patterns and may rely

more heavily on patterns of variation at the level of stereotypes. These stereotypes could be derived from experience with imitated or mock varieties of the target variety (e.g. in films or by comedians), but they may also reflect experience with meta-commentary about the variety (e.g. that Boston speakers drop their [ɹ]s or that Chinese has tones).

The task of determining authenticity is analogous to the lexical decision task described above for determining semantic priming. Listeners are not asked to provide an explicit label for a voice such as “Chinese” or “not Chinese,” but are asked to identify the voice as “authentic” or “not authentic”. The distinction may seem subtle, but it is crucial. Listeners should be able to draw on the same linguistic and social ideologies they might use outside the laboratory when sizing up a new interlocutor or enjoying an actor’s performance. Neuhauser and Simpson (2007) argue that non-native listeners and accent imitators share a surprisingly uniform cognitive prototype of what features an imitated variety should have, while true non-native speakers both produce and anticipate patterns which defy these common expectations. For example, hearing an authentic variety will present listeners with accurate variety-specific patterns of co-articulation (Beddor *et al.* 2002). Listeners with more extensive experience with the target variety should be sensitive to these patterns. Listeners with less experience, on the other hand, have no reason to expect or to prefer these patterns that exist below the level of meta-linguistic commentary. Indeed, there is some evidence that less experienced listeners will actually be more drawn to the phonetic implementation of the imitated accent. Neuhauser and Simpson (2007), for example, 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.

The German finding points to another interaction between awareness and experience that is crucial for modeling the likely exemplar representations of a listener. Specifically, some listeners will have more detailed social category representations than others. Additionally, some listeners may be capable of forming more accurate linkages between acoustic signals and social category representations (this continuum of accuracy is analogous to Preston’s “detail”). In principle, these two levels of representation and encoding are independent so that one listener may be very good at linking fine phonetic detail to rather broad social categories (e.g. an acting coach), while another listener may have quite detailed social categories, but difficulty linking fine phonetic detail to those categories (e.g. Milroy and McClenaghan 1977). This claim is analogous to the observation made by Wells (1982) and cited in Agha (2003) that a working class accent may sound merely British to a Chicagoan, English to a Glaswegian, Northern to a Southerner, Liverpoolian to a Northerner, and working class to a Liverpoolian, except that this increasing complexity is not accurately described as a simple fact of one’s geography. There is no reason

that a Chicagoan should not, through eager attention, have quite accurate social categories for English people and accurate linkages from sociolinguistic variables to those categories. Finally, there is no reason to expect that the sheer amount of experience a Liverpudlian may have with local varieties of English will necessarily result in rich social categories and accurate linkages of fine phonetic detail to those categories. These linkages are a function of awareness of both phonetic detail and detailed social categories, attention to acoustic and contextual cues, and encoding of these cues so that the exemplar dynamics resulting from a particular degree of exposure to a variety is likely to vary from individual to individual.

In the present chapter, more experienced listeners are more accurate at both selecting the authentically Chinese-accented voice and at rejecting the imitated variety. However, less experienced listeners do perform above chance levels on authenticity detection. This finding, together with Neuhauser and Simpson's finding about German listener preferences for imitated varieties, leads to the question of what it is listeners are attending to in these imitated varieties. The second half of this chapter reports a laboratory investigation of actors' imitations of Chinese-accented English. This investigation is intended to help us to begin to understand the constellation of linguistic cues native English listeners use to perceive a particular voice as Chinese or, as we will see, as 'Asian'. The acoustic correlates of 'Chinese-ness' for less experienced listeners will be explored through a production study with a small group of American-English-speaking actors attempting to produce a percept of 'Chinese' for an American, perhaps monolingual, English-speaking audience through imitation of Chinese-accented English.

Actors were hired to perform the imitated variety rather than eliciting folk linguistic imitations (cf. Brunner 2010). This decision is motivated partially by expediency – with the assumption that actors will more successfully and consistently perform the imitated accent. More important, though, are the theoretical motivations. The actors choose to imitate some features and to ignore others with the goal of invoking a 'Chinese' character type, as defined in the expectations of their audience. It is these percepts, and the features that encourage them, that we are interested in understanding. It is presumed that perception of less successful or less consistent imitated speech will be perceived via the same mechanism and will activate the same representations, only less clearly.

### **Quantifying Listeners' Experience**

Quantifying listener experience with, and expectations about, language is by no means unique to questions of social experience nor indeed even to speech perception. In diverse linguistic or psychological experiments, it is often



necessary to estimate, for example, the frequency of a particular lexical item in listeners' experience outside the laboratory.

The normal practice for estimating listeners' exposure to lexical frequency is to calculate these frequencies from an established corpus such as Kucera and Francis (1967) or Baayen *et al.* (1993). Although both corpora are quite dated and drawn from a mixture of print and spoken sources that are unlikely to represent the statistical patterns experienced by modern participants (Balota *et al.* 2007), these data provide an expedient and, more importantly, standard surrogate for listener experience with linguistic forms. The following section argues for the particular importance of quantifying not only listener experience, but also listeners' language ideologies in socially cued perception research.

Work in socially cued speech perception critically relies on an understanding of the identity of the listeners – their linguistic experience, how they relate to the experimenter, how they relate to the variety being studied, etc. The need to quantify listener experience is therefore likely greater than in other linguistic and psycholinguistic experimentation. Here, we have all of the same questions of frequency and patterning of linguistic forms, but with the added recognition that these forms will differ systematically by listener and social context.

It is pragmatically useful to conceive of identity as a monolithic, fixed property of an individual, but this is also a massive simplification. Identity is much better understood as a dynamic, context-sensitive construct in which interlocutors manipulate and interpret indexical forms to define their roles in a particular interaction (Bucholtz and Hall 2010). Irvine (1989) refers to “a diversity on the linguistic plane that indexes a social diversity” and recent work in sociolinguistics and linguistic anthropology has demonstrated that speakers and listeners are aware of, and exercise situational control over, these diversities. Depending on context, speakers will invoke different constellations of indexical linguistic forms – different registers – to convey the same denotational or referential meaning (Silverstein 2003). In other words, not only does one speak differently in a job interview than one speaks in casual conversation, but listeners are aware of and expect these uses of appropriate registers. Violating these expectations is possible, but this is a meaningful act.

Babel (2010) reports the use of Spanish/Quechua contact features among speakers in one Andean village. Given local ideologies that associate Quechua use with informal, rural speech, it is unsurprising that Quechua-influenced contact features in this variety of Andean Spanish are more commonly used in informal conversation than in interview or meeting contexts. However, these features are also invoked in more formal contexts as indices of the speaker's authenticity, to create intimacy, to mark one's affiliation with a particular political group, and sometimes several of these social meanings simultaneously. Individual linguistic forms may be linked to particular social meanings, but both the linkage and the meaning are highly context-dependent.

For the experimentalist, then, it is worth bearing in mind that performance on a task intended to quantify listener experience will be shaped by the formal, experimental context, by listener ideologies about the details of the language being used, and by listener ideologies about the purported speaker.

For experimental investigation of exemplar dynamics, it would be ideal to have a range of background information about each participant that it is difficult to conceive of collecting either behaviorally or through self-report. For a study like Rubin (1992) or McGowan (2015) in which listeners see an Asian or Caucasian face while listening to stimuli, this background information includes such variables as the frequency and intensity of interaction with Asian interlocutors, the probability of an Asian face accompanying a non-native English accent in the listener's experience, the distribution of facial features and L1 languages and their combination in the listeners' experience with speakers, how the listener construes the phonetic features under investigation to create meaning, how the listener self-identifies linguistically, and – critically – awareness of and attention to any acoustic cues to Asian-ness. Every aspect of stimulus presentation is potentially open to influence from listener experience and ideology.

This depth and breadth of understanding is probably impossible to achieve in the laboratory. The standard solution is to ask participants to complete a language history survey – typically in the laboratory or when registering for a subject pool. Survey instruments vary, but they generally request such information as the participants' native language(s), languages spoken at home, languages studied, places the participant has lived, etc. Experience can be inferred from such a survey instrument, but questions can necessarily only access knowledge above the level of conscious awareness.

The approach to listener experience quantification taken here adapts a task from forensic phonetics (Neuhauser and Simpson 2007) and is an attempt to assess participants' ability to correctly identify an authentic Chinese accent from a set of distractor accents. The assumption, again, is that more experienced listeners will, in the general case, have greater sensitivity to fine phonetic detail consistent with an authentic, rather than imitated, Chinese-English accent; at the same time, that less experienced listeners will be more drawn to an imitated variety. Both listener populations will apply what knowledge and stereotypes they have available to guide speech perception, but their differing levels of experience and their differing relationships to 'Chinese-ness' should be discernible via the authenticity detection task.

However, it must be noted that the generalization "Chinese accent" is so broad as to be almost offensive. The so-called "dialects" of Chinese include six separate language phyla: Sino-Tibetan, Austro-Tai, Austronesian, Altaic, Austro-Asiatic, and Indo-European. Many of these dialects are not mutually intelligible, with listener subjective ratings of mutual intelligibility closely

matching performance on cross-dialectal semantic classification and speech-in-noise perception tasks (Tang and van Heuven 2009). This suggests that, although Chinese L2 English speakers may all *also* command Mandarin, or Standard, Chinese, Mandarin is itself likely to be an L2 language or spoken with the accent of a regional dialect. Additionally, different non-native English speakers from China have acquired different target Englishes. Until quite recently it has been the norm for Chinese students of English to target RP-accented British English as their normative model for acquisition. Increasingly, though, students target American English or even a contact variety known as “China English” (Qiong 2004).

### **Experiment 1: Authenticity Detection Task**

Experiment 1 is an accent authenticity detection task using the *yes/no* detection paradigm. The listener is presented with a single stimulus recording per trial and must press one button on a response box if the stimulus sounds like authentic Chinese-accented English and another button if the stimulus sounds like some other form of accented English. This experiment was designed to measure listeners’ ability to correctly detect an authentic Chinese accent among a collection of accents that include authentic Chinese and imitated Chinese as critical stimuli, together with a set of filler accents.

One goal of this experiment was to quantify the extent to which listeners with little or no experience listening to native speakers of a target variety nevertheless use social knowledge in a systematic way during perception and to compare this performance to that of more experienced listeners. It is hoped that populations of more and less experienced listeners can be identified and their experience quantified using this method. This categorization should be particularly useful when the variety or variable in question is below the level of available, conscious awareness. It should then be possible to test the attribution of socially cued effects in speech perception to stored episodic traces of linguistic experience linked to social knowledge.

## **Methods**

### *Stimuli*

Stimulus materials consisted of the eight sentence types listed in Table 2.1. These were all English recordings spoken by two native speakers of Mandarin Chinese and one native speaker each of Korean, Turkish, and Macedonian, all drawn from the Wildcat Corpus (Van Engen *et al.* 2010). The Wildcat Corpus includes individual words, the “Stella” passage from the Speech Accent Archive at George Mason University, the “North Wind and the Sun” from

Table 2.1 *Sentences Used in Experiment 1*


---



---

She made the bed.
Bob wore a watch on his wrist.
Dad talked about the bomb.
I wear my hat on my head.
The color of a lemon is yellow.
A racecar can go very fast.
He looked at the sleeves.
Please call Stella.

---



---

the IPA Handbook, high and low predictability sentences, and unscripted recordings from a map task. The sentence recordings from this experiment were drawn from the scripted passages and sentence recordings. These stimulus recordings were augmented with recordings of two monolingual English speakers – not trained actors – performing imitated Chinese accents.

In general, the accuracy of the imitated Chinese was poor, but as in Neuhauser and Simpson (2007) speaker manipulations were consistent. Native speakers of Midwestern American English from Michigan were asked to read, with an imitated Chinese accent, the same texts recorded by authentic Chinese-accented speakers for the Wildcat Corpus. These speakers were not presented with a model accent to imitate. The voices selected for inclusion in the study imitated the authentic backing of interdental fricatives (/ð/ → [z] and /θ/ → [s]) and the stereotypical feature /ɹ/ → [l] that is rarely, if ever, found in authentic Chinese-accented English. The American-English speakers, from southeastern Michigan, who produced the imitated Chinese consistently produced post-vocalic /ɹ/, while the authentic Chinese-accented speakers did not.

Figure 2.1 shows a spectrogram of a sample token of authentic Chinese-accented English. This male speaker has produced the word *racecar* as [ɹeɪsk<sup>h</sup>aː]. The word-final vowel is rhoticized for nearly its entire duration with no audible consonantal articulation. There is a pitch contour on this syllable similar to the Mandarin Chinese third tone with its characteristic dip and rise.

Figure 2.2 shows a spectrogram of a sample recording of imitated Chinese-accented English. This male speaker has produced the word *racecar* as [ɹeɪsk<sup>h</sup>ɑː]. This speaker generally replaced /ɹ/ in non-post-vocalic positions with [l]; however, in this particular token there is visible and audible friction. The post-vocalic /ɹ/ is clearly visible and audible over the last 71 ms of the token and the vowel is audibly rhoticized for 50 ms (six glottal pulses) prior to the consonantal articulation. That the imitated Chinese speakers consistently produced post-vocalic central approximants is initially surprising. The lack of post-vocalic /ɹ/ is a stereotypical feature of Chinese-accented English and one that most of the professional actors in the next section also imitated.

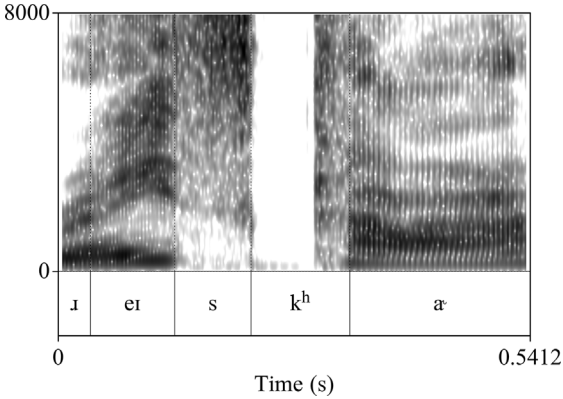


Figure 2.1 Spectrogram of Male Authentic Chinese Speaker Producing *Racecar*  
Post-vocalic /ɹ/ is clearly absent in the spectrogram.

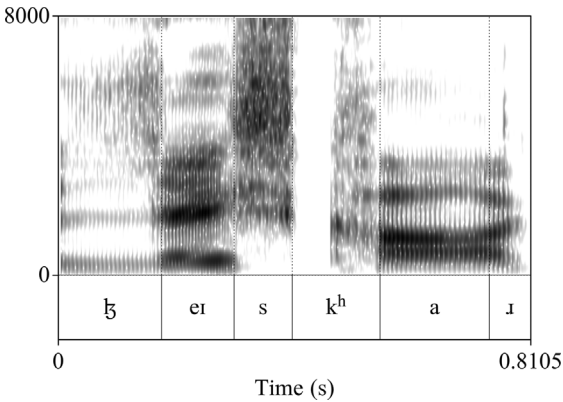


Figure 2.2 Spectrogram of Male Imitated Chinese Speaker Producing *Racecar*  
In this imitation, initial /ɹ/ has been replaced with a voiced alveolar lateral fricative. Post-vocalic /ɹ/ is clearly visible in the spectrogram.

## Participants

Eighty-seven undergraduate students participated at one of two experiment sites. All participants provided self-reported experience ratings (more experienced versus less experienced with Chinese-accented English) on a language history survey. 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 statements regarding having experience listening to Chinese-accented English; having close friends who speak Chinese as a first language; 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” had a mean age of 19, had lived 98 percent of their lives in the United States, were predominantly born in Michigan (51 percent) or New York (18 percent), reported having no friends or family members who spoke Chinese as a first language, and on average claimed not to have a clear idea of 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” had a mean age of 22, had lived 82.3 percent of their lives in the United States, were predominantly born in California (55 percent) or China (38 percent), 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 about 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 the margin, “unless you’re Asian!”.

The original intention had been to run the entire experiment at the University of Michigan research site, but locating participants who self-reported as “more experienced” at the Michigan site proved unsuccessful. For this reason, a second site was added at the University of California, Berkeley, where such a population was more readily identified.

### **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. Five participants were identified for exclusion prior to analysis for reporting experience with Mandarin Chinese – either through language study or, in four cases, for being bilingual or Heritage speakers. These participants will be included in the correspondence analysis to test the authenticity detection task’s ability to accurately exclude non-representative participants, but these participants will be excluded from other statistical and visual data analysis. One participant was excluded for using Facebook and sending text messages on his smartphone during the experimental session. One additional participant was excluded from the data analysis

for struggling to remain awake during the experiment and reporting the task as extremely difficult. Three data files were lost due to experimenter error.

### **Berkeley Listeners**

Heritage speakers with little or no proficiency in Mandarin were selected as a target population. This selection was intended to avoid the complications of interpreting the results of truly bilingual speakers for what is essentially an English language task.

Thirty Heritage Mandarin-speaking undergraduate students from the University of California, Berkeley participated in exchange for an incentive of \$15.00 per participant. Two participants were removed prior to any analysis: one L1 speaker of Mandarin who had misunderstood the flier, and a second individual who misrepresented his identity. As with the excluded listeners from the Michigan group, these participants will be included in the correspondence analysis, but excluded from other statistical and visual data analysis. Time constraints limited the number of participants who could be engaged in the study at this site.

### **Procedure**

All listeners used Apple Macbook Computers (model 4,1; late 2008). Testing of Michigan listeners took place in an IAC sound-attenuated booth in the University of Michigan Phonetics Lab; stimuli were presented over AKG K271 mkII headphones. Responses were entered via Cedrus RB-620 low-latency response boxes with serial to USB adaptors.

UC Berkeley listeners used the same computers and software as at the Michigan site. However, these testing sessions took place in the phonology laboratory at the University of California, Berkeley. This is a quiet space dedicated to speech perception experiments, but is not a sound-attenuated booth. AKG k240 headphones and Cedrus RB-730 low-latency USB response boxes replaced the headphones and response boxes used at Michigan.

Stimuli were presented using Superlab stimulus presentation software version 4.0.8. Volume was set at a comfortable listening level. Listeners indicated their responses via button box. Listeners were instructed to press one button if the voice they heard had an authentic Chinese accent and another if the accent was not authentic Chinese. The target sentences were presented on-screen from the onset of the recording playback until the subject submitted a response. Listeners were informed that the voices would include a range of different non-native English accents, including Chinese, imitated Chinese, Korean, Turkish, and Macedonian. It was not possible to change responses or to hear recordings more than once. Listeners were encouraged to rest after

each block and there were enforced breaks at the halfway point. Participants responded to eight sentences produced by seven voices in each of six blocks for 336 responses per participant.

All participants in this experiment had just completed the transcription-in-noise task reported in McGowan (2015). No voices or stimuli were repeated from that experiment.

## Predictions

The use of imitated Chinese accents was inspired by Neuhauser and Simpson (2007), who found that German monolingual speakers were more likely to identify German imitations of French and American accents as authentic than they were to correctly discriminate true non-native accents. I hypothesized that native listeners must be drawing on language ideologies concerning foreignness in general and the target non-native accents in particular when making discrimination judgments. This hypothesis, if supported, would have implications for research in socially cued speech perception that has appealed to stored episodic traces to explain behavioral results.

It is difficult to imagine a means of differentiating between listener knowledge gained through real communicative experience with a language variety and listener knowledge of linguistic stereotypes (again, in the sense of Labov 1972) gained through exposure to imitations of that variety or occasional brief exposure in the media. However, the Neuhauser and Simpson (2007) result suggests one possibility. If less experienced and more experienced listeners are drawing on both qualitatively and quantitatively different forms of knowledge – in terms of both amount of experience and accuracy of linking a Chinese-accented voice to a “Chinese” social category – when detecting an authentic Chinese accent, then they should be differentially drawn to authentic and imitated stimuli.

## Results

### *Proportion “Yes” Responses*

Figure 2.3 shows proportional “yes” responses to each non-native accent by experience level. More experienced listeners are more likely to respond “yes” to an authentic Chinese voice. More experienced listeners are also more likely than less experienced listeners to respond “yes” to an authentic Chinese accent. Less experienced listeners, by contrast, appear to be more likely than more experienced listeners to identify an imitated Chinese accent as “authentic.” This pattern of responses suggests that more and less experienced listeners are employing different strategies when deciding whether a particular voice is “authentic.”



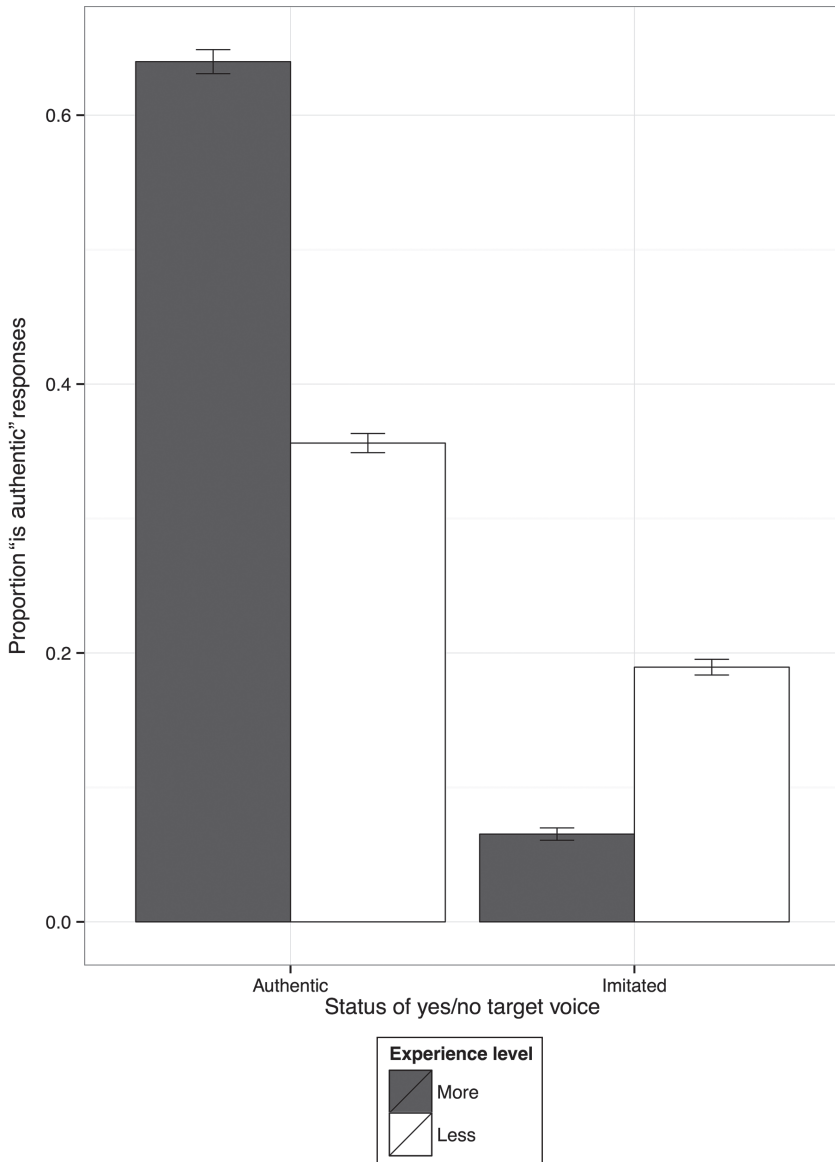


Figure 2.3 Proportion “Yes” Responses by Accent and Experience

The yes/no responses were then analyzed using the open source statistical package R (R Development Core Team 2014). The data were modeled with linear mixed-effects models, as implemented in the lme4 R package (Bates *et al.* 2014). Categorical variables were sum-coded to allow the interpretation

Table 2.2 *Fixed Effects with Coefficients and P-Values for “Yes” Responses by Accent and Experience*  
*Reference levels: accent – Chinese; experience – Experienced*

(Ref. level: experienced; Chinese)	Coef $\beta$	SE( $\beta$ )	z	p
(Intercept)	1.21	0.26	4.68	<.001
Accent	-1.92	0.05	-37.54	<.001
Experience	-0.02	0.29	-0.08	0.94
Accent: Experience	1.43	0.06	24.89	<.001

of any 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 *et al.* 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 *et al.* 2013).

The yes/no responses were analyzed with Subject and Item included as random effects with both random intercepts and random slopes for Experience. The dependent measure in this model was whether the participant responded “yes” to the stimulus (“yes” response regardless of accuracy is an indicator of the listener’s belief that the stimulus is authentic Chinese). Accent and Experience were included as fixed effects. A fuller model including the interaction of Accent and Experience provided a better fit for the data than a reduced model with no interaction term ( $\chi^2(1) = 792.08$ ;  $p < 0.001$ ). Factor levels, coefficients, standard error, z-score, and p-values for each level of these factors and interaction are reported in Table 2.2.

There is a significant main effect of Accent ( $\beta = -1.92$ ,  $p < 0.001$ ). Surprisingly, given the apparent differences in Figure 2.3, there is not a main effect of Experience. However, the interaction of Accent and Experience is significant, demonstrating that more and less experienced listeners are drawn differentially to the authentic and imitated Chinese accents.

### Accuracy

Table 2.3 shows percentage accuracy on the authenticity detection task by target Accent and self-reported Experience level. As anticipated, more experienced listeners appear to be more accurate when responding to either authentic Chinese or to imitated Chinese stimuli. However, accuracy results do not, on their own, necessarily reveal the listeners’ ability to detect a signal such as the Chinese accent in this experiment. A listener hoping to have perfect recall on

Table 2.3 *Percent Correct Responses by Accent and Experience*

Experience	Authentic Chinese	Imitated Chinese
More experienced	64.0%	93.5%
Less experienced	35.6%	81.1%

Table 2.4 *Signal Detection Results Authenticity Detection Task*

Experience	Hit rate	False alarm rate	$d'$	$c$
More experienced	0.64	0.06	1.87	0.58
Less experienced	0.36	0.19	0.51	0.62

the Chinese-identification task could, for example, simply press the “yes” button in response to each stimulus item. Overall performance would be poor, but accuracy to the Chinese stimuli would be perfect.

A measure of response sensitivity from signal detection theory,  $d'$ , represents the distance between a listener’s ability to maximize hit rate (correct identifications) and minimize false rejections. Table 2.4 reports hit and false alarm rates in these results, together with  $d'$  and criterion  $c$  scores. The question addressed by these metrics is the extent to which listeners are correctly identifying authentic Chinese and rejecting other accents. The much higher  $d'$  score for more experienced listeners suggests that, as predicted, these listeners are much more sensitive to the differences between authentic and imitated Chinese-accented English.

The criterion measure, or  $c$ , is a measure of response bias that attempts to model the decision criterion chosen by listeners when completing a task.

$$d' = z(H) - z(F) \quad (1)$$

$$c = -0.5[z(H) + z(F)] \quad (2)$$

H represents the hit rate: correct “yes” responses divided by possible “yes” responses (equivalent to recall in information retrieval). F represents the false alarm rate: incorrect “no” responses divided by the number of potentially correct “no” responses).  $z()$  is a z-transform function (taking probabilities and returning z-scores). Positive  $c$  scores correspond to a tendency to respond “no” during the task. Both groups of listeners are biased to respond “no,” but experienced listeners much more weakly so. If  $c = 0$ , the listener is unbiased; less experienced listeners have a slightly stronger “no” bias ( $c = 0.62$ ) than more experienced listeners ( $c = 0.58$ ). This bias to respond “no” is likely attributable to a weakness in the task’s design. With three filler accents and two

imitated Chinese accents, only 25 percent of the trials required a legitimate “yes” response. It is likely that a replication of this study, which removed the probably unnecessary filler accents, would not only make the task shorter and easier to administer, but also obtain even stronger d-prime results and, therefore, a more accurate predictor of group membership.

### Clustering

Figures 2.4 and 2.5 present a visualization of a correspondence analysis of the authenticity detection data, including both critical and filler trials. Correspondence analysis is an unsupervised clustering technique. From a contingency table of “yes” responses by participants to each level of the language factor, two

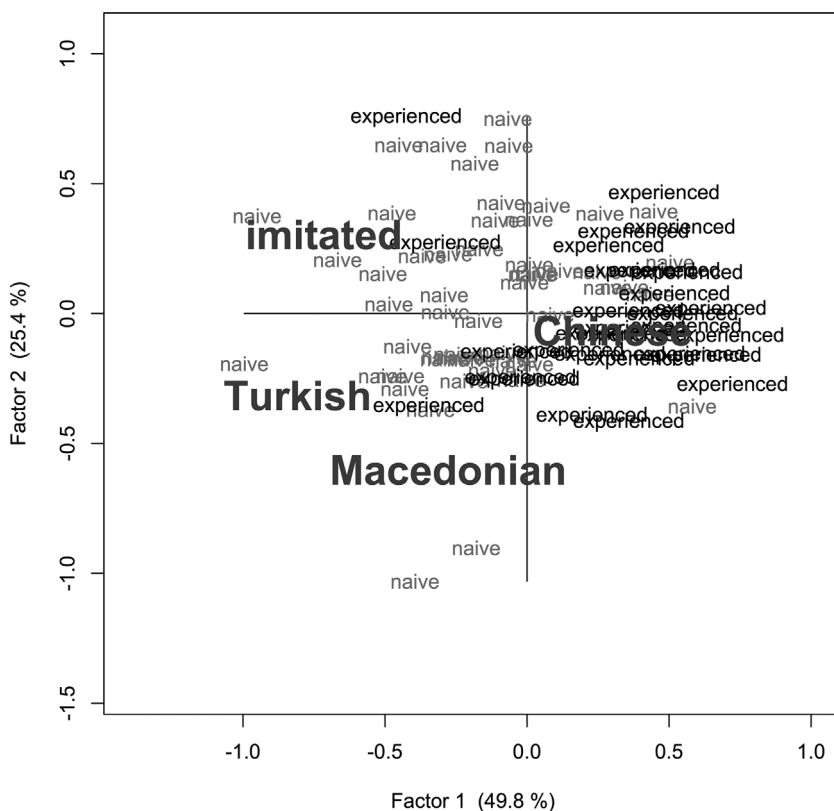


Figure 2.4 Correspondence Analysis of More and Less Experienced Listeners. More experienced listeners cluster tightly around the authentic Chinese target, while naive (less experienced) listener responses are more diffuse.

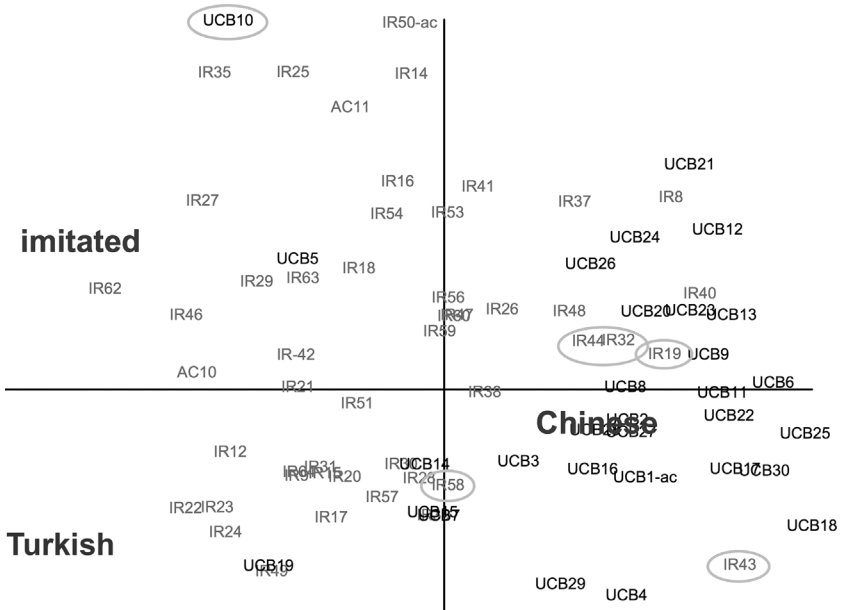


Figure 2.5 Correspondence Analysis of More and Less Experienced Listeners Cropped and Zoomed to Highlight Participant ID Detail  
 Circled participant IDs were independently excluded prior to further data analysis; “ucb” indicates more experienced listeners and all others are from the less experienced group.

separate square distance matrices are calculated: a row-by-row matrix (in this case, distances between participants) and a column-by-column matrix (distances between languages). The software used here, Baayen’s LanguageR package for the R open source statistical environment (R Development Core Team 2014), uses a chi-squared distance metric. Like principal components analysis does for real-valued data, correspondence analysis provides a low-dimensionality map of both rows and columns in a contingency table (Baayen 2008). “Factor 1” on the x-axis represents the most informative column, authentic Chinese, with an eigenvalue rate of 0.4984 or 49.84 percent of the variance in the table. “Factor 2” on the y-axis represents the second most informative column, imitated Chinese, with an eigenvalue rate of 0.2538 or 25.38 percent of the variance in the table. This two-dimensional visualization of the data captures roughly 75.2 percent of the variance in the table; Korean contributed virtually no explanatory power to the map and has dropped out of the visualization.

Intuitively from Figure 2.4 we can see that the more experienced Heritage Mandarin participants from the University of California, Berkeley have, for the

most part, clustered tightly around the Chinese label. This suggests that, as predicted, these listeners were more attracted to Chinese for responses of “authentic Chinese” than to any other language. The clustering of less experienced (here rendered as “naïve” for visual differentiation) monolingual English participants from the University of Michigan is much more diffuse. They appear to be attracted to both the imitated and authentic Chinese languages for “yes” responses, with neither cluster being a particularly good predictor of less experience.

Interestingly, all but one of the participants who were independently excluded from data analysis are outliers on this plot. Figure 2.5 is a zoomed and cropped view with the excluded participants circled. Participant UCB10 was excluded from the experienced data set for misrepresenting his identity and, reassuringly, is among the least experienced of the less experienced participants in terms of attraction to the imitated Chinese voices. Participants IR19, IR32, IR43, IR44, and IR58 were excluded from the less experienced data set for self-reporting extensive or Heritage experience with Mandarin-accented English. Of these, only IR58 does not clearly cluster with the experienced participants.

### **Reaction Time**

It was predicted that experienced listeners should have lower reaction time latencies overall. The question, after all, is whether the voice is authentic Chinese and these listeners have copious experience with this variety to draw upon. This prediction was not upheld. There were no predictions regarding the relative time required to respond to Authentic versus Imitated Chinese-accented stimuli, however, as shown in Figure 2.6, responses to the Authentic stimuli were 418 ms faster, on average, than responses to the Imitated stimuli. Less experienced listeners were also, on average, 155 ms faster than more experienced listeners when responding. Reaction times longer than two standard deviations above the grand mean were excluded from analysis. Even after the log transform there remain large numbers of slow outliers (visible as black circles on Figure 2.6) and this is true across listeners, across accents, and across experience levels.

### **Discussion**

The question posed in the authenticity detection task was whether ability to discriminate an authentic from an imitated Chinese voice might be a good predictor of listener experience level. The first prediction was that more experienced listeners would more accurately identify the authentic variety. This prediction was upheld. More experienced listeners are better able to identify

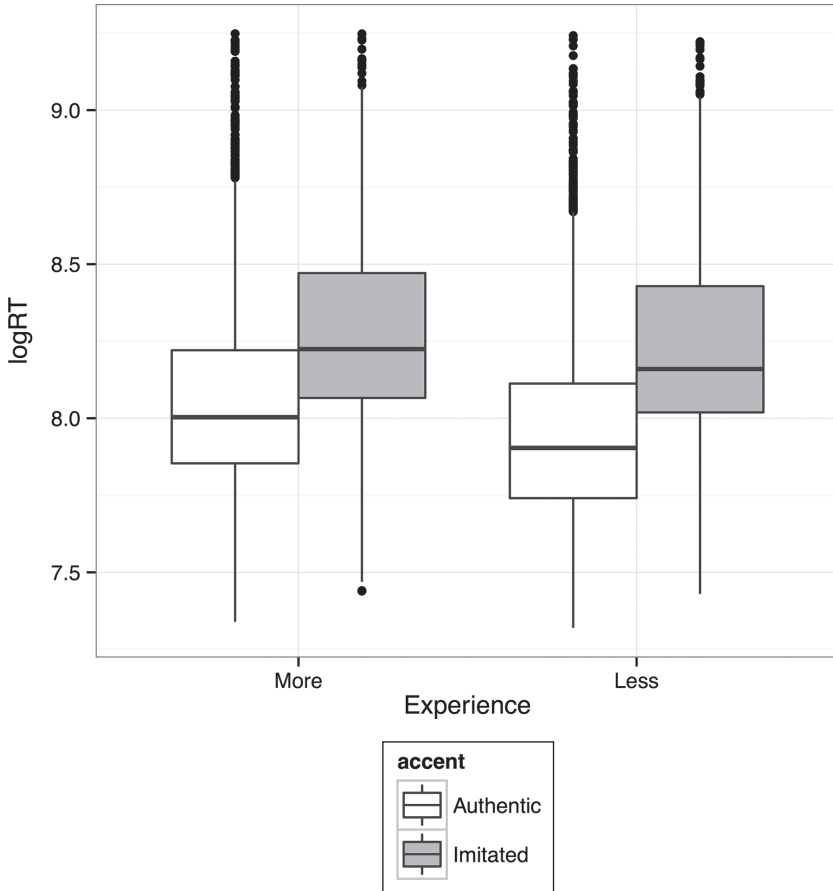


Figure 2.6 Log Transformed Reaction Times by Accent and Experience

the authentic variety (as shown by the d-prime analysis) and less likely to be charmed by the imitation, as revealed in the analysis of “yes” responses and the correspondence analysis. This result is hypothesized to be due to three factors: the accumulation of detailed episodic traces of linguistic experience; a clearer mental representation of a “Chinese” social category; and more accurate linkages between fine phonetic detail and the social category label of “Chinese.”

The second prediction was that less experienced listeners would be more drawn to the imitated variety. This prediction was also upheld with less experienced listeners being more drawn to the imitated variety, as revealed by the “yes” responses analysis, and less able to distinguish authenticity as shown by d-prime. Interestingly, while less experienced listeners are less

accurate than experienced listeners at identifying an authentic Chinese voice, their performance is above chance. They are successfully drawing on *some* kind of knowledge to perform this task and this knowledge is unlikely to be a rich cloud of Chinese-accented English exemplars with accurate linkages to a well-defined “Chinese” social category. It seems likelier, given the responses on the language history survey, that these listeners are drawing on culturally available stereotypes of how Chinese-accented English sounds. This suggestion is, again, consistent with Neuhauser and Simpson’s findings for German listeners hearing French and American accents.

It seems reasonable to infer, then, that more and less experienced listeners are drawing not merely on different *amounts* of knowledge, but on different *forms* of knowledge when detecting an authentic Chinese accent. This finding does not refute exemplar models, but it does suggest a need for a more nuanced understanding of the knowledge listeners use to structure speech sensory data. This more nuanced view should highlight the role of listener ideologies and listener awareness of social categories during perception. The more experienced listeners could well be drawing primarily on stored episodic traces of experience with authentic Chinese-accented English, while less experienced listeners draw on stored episodic traces of comedians and actors imitating the accent. However, it should be noted that even the most experienced listener will also anticipate stereotypical features of a variety and that even in these listeners stereotype and experience must therefore interact in perception.

So, too, it should not be assumed that less experienced listeners’ success on this task implies that they conceptualize “Chinese” in the same richly detailed way that the more experienced Heritage Mandarin listeners do. White Americans typically conflate all ethnicities and nationalities from East Asia into a single pan-Asian, pan-ethnic group, who, while believed to be a “model minority” are nevertheless conceptualized as perpetually foreign and essentially unassimilable (Espiritu 1992; Wong *et al.* 1998; Iwamoto and Liu 2010). For these listeners, detail about sociolinguistic features is likely to be low and the social categories to which these features are linked is likely to be coarse-grained, combining aspects of Chinese-accented English with, for example, Japanese- and Korean-accented English. These issues will be explored further in the next study, which looks at the production of imitated Chinese by less experienced, American actors.

### **Usefulness as a Replacement for, or Supplement to, Self-Reported Experience**

An additional goal of the authenticity detection task was to explore the usefulness of this task as a means of directly estimating participants’ experience level with authentic Chinese-accented English. The *yes/no* task presented here is



extremely easy to build and administer for any target variety, requiring only a fairly small set of authentic and imitated language recordings.

A task of this sort should be particularly helpful in gauging listeners' experience with language varieties for which the less experienced and more experienced populations are not so easily identified as Mandarin-accented English. It could also be helpful in assessing listeners' experience with varieties that they may have ideological reasons to disavow knowledge of (e.g. middle class African American students might wish to distance themselves from knowledge of African American Vernacular English (AAVE) – particularly in a formal context). Finally, this tool may be of use in the investigation of varieties or sociolinguistic variables that, like lexical frequency or compensation for co-articulation, are entirely below the level of listeners' conscious awareness.

This section closes with an important caution about the interpretation of the present result. Although every attempt was made to keep the experiment as consistent as possible despite the change of venues, the less experienced and more experienced listeners were, in retrospect, fundamentally not performing the same task. Although they used the same computers, were given the same instructions, heard the same stimuli, saw the same sentences, and performed the same physical task, more experienced listeners were inescapably aware of having been recruited precisely because they were Heritage speakers of Mandarin Chinese. The less experienced listeners were simply asked to identify the authenticity of a non-native accent. More experienced listeners, by virtue of their background, are not only trying to identify the authenticity of an accent, but, in a very real sense, are also striving to demonstrate and possibly even defend their own authenticity. I believe this difference alone accounts for experienced listeners' somewhat slower reaction times. Future uses of this technique will need to be more careful about keeping experienced participants naive to the role of their own experience in the experiment.

### **Production of Imitated Chinese**

It is difficult to discern from the results of the first experiment which features listeners were attending to when making their determination of authenticity. This task is made more difficult by the interrelatedness of the various cues that might be active at any given moment: speaking rate, pitch variability, and segmental alternations are each likely to vary in any particular word or even syllable that might be selected for presentation to a listener. This problem is, of course, not unique to the present study and numerous solutions have been used in the past to isolate the informativeness of particular acoustic features (e.g. synthesis, modification of natural speech, eye tracking, masking, etc.). However, because the subject of the present investigation is listeners'

awareness and control of socially informative variation, this study moves from perception to production. Specifically, professional actors who are themselves native Michigan-English speakers and who self-identified as inexperienced with authentic Chinese-accented English were hired to perform scripted materials with an imitated Chinese accent.

It is hypothesized that, in order to create a percept of *Chinese* for an English-speaking audience, these actors will perform the same stereotypical features that less experienced listeners drew upon in Experiment 1 to make their identification decisions. Actors use imperfect imitations of “dialects” to create characters. Successful actors achieve this quickly and efficiently. This experiment is, again, inspired by Neuhauser and Simpson (2007), whose native German listeners were more likely to identify a French or American accent performed by a fellow German as authentic than truly authentic French- or American-accented speech. Actors train in many different vocal and speech traditions, but will generally only alter a specific subset of their productions which are intended to be the characteristic speech sounds of the target dialect or variety. Blumenfeld (2002), for example, provides the aspiring actor with a wide range of accents and lists of the “most important sounds” required to perform the target accent. By manipulating these stereotypical aspects of their speech, though, the actor constructs with the listener an idealized representation of the target variety. Authenticity, for a less experienced listener, may well be measured not only by how consistently the actor makes these substitutions and how well aligned these substitutions are with the listener’s own stereotypical expectations, but also, in some sense, by how comfortable and familiar the non-stereotypical aspects of the voice remain.

### Chinese-Accented English

To establish a baseline for the particular features the actors are likely to imitate (and thus guide the phonetic analysis), it may be useful to compare the perceptions of Chinese-accented English made by linguists to language ideologies held by less experienced listeners.

I interviewed a linguist employed at the University of Michigan’s English Language Institute, which offers English language instruction, counseling, and testing to members of the University of Michigan community. This linguist provided the following list of features that typically prove challenging for L1 Mandarin speakers seeking English proficiency through the Institute.

<b>/b/ /d/ /g/</b>	devoiced in final position (lab/lap, mob/mop, bed/bet, mad/mat, hard/hot, lag/lack, dog/dock)
<b>/v/</b>	realized as [w] or sometimes [b] (vision, vet/wet, vine/wine, provide, university, overseas, involve)

/θ/ /ð/	tongue tip is closer to the alveolar ridge than to a dental or interdental articulation resulting in [s] or [z] voiceless: (thing/sing, thank/sank, faith/face, math/mass, method) and voiced (the, mother, either, weather, etc.)
/ð/	for some speakers (this may be regional or related to an L1 other than Mandarin), /ð/ is realized as /d/
/ʃ/	more palatalized than English [ʃ]
/ʒ/	realized as [w], [j] depending on vowel context or as a voiced palato-alveolar affricate [dʒ] (e.g. usual [juwəl], vision [viʒən])
/dʒ/	occasionally realized as [ʒ] word-medially (virgin/version, ledger/leisure)
[r]	realized as [t] / [d] (water, party, kitty, city, ladder)
[ʔ]	realized as [t] / [d] (mountain, kitten, fantasy)
/m/	sometimes alveolar in final position (some/sun, rum/run). Rare?
/n/	consonant deleted (nasalization seems to be moved to vowel) (untie, inside, romantic, human, tradition)
/ŋ/	realized as [n] or deleted (with $\tilde{V}$ ) (thing, long, length)
/l/	realized as [oʊ], [w], or deleted word finally (cold/code, fault/fought, dull/dough, call/caw, all/awe, social, people)
/ɹ/	deleted word finally for some speakers (not Beijing) (far, order, error, turn/ton, bird/bud, work/walk, mark/mock)
<b>consonant clusters</b>	simplified; often partially or wholly deleted – even across syllable contact boundaries (coul(d), ou(t)side, pra(c)tice, remi(nd))

### Folk Observations

Lindemann (2005) asked American-English listeners to label maps with descriptions of English as spoken by international students. These students provided largely negative evaluations of Chinese-accented English with, Lindemann notes, “a surprising amount of agreement” in their qualitative descriptions of the salient features of Chinese-accented English. Lindemann’s data is summarized in Table 2.5.

Lindemann’s respondents chiefly focus on Chinese speakers’ confusion of /ɹ/ and /l/ – a feature notably absent from the linguists’ description.

### Method

Five English-speaking actors, all natives of southeastern Michigan, were hired to perform a set of scripted materials. These materials included the Stella passage, “The North Wind and the Sun,” and the same thirty pairs of high and low predictability sentences from Bradlow and Alexander (2007) used in Experiment 1.

Table 2.5 *Folk Features of Chinese-Accented English (from Lindemann (2005))*

---



---

Speak quickly
Pronounce L's as R's.
Voices rise when cursing
Choppy speech
High toned/high pitched English
Missing verbs (copula)
Forget to add plural "-s"
Difficult to understand

---



---

All five actors self-identified as inexperienced with Chinese-accented English – apart from their preparation for this performance. Each actor was asked to read the scripted materials twice. First they read in their normal speaking voice (which, without exception, was a performed stage version of their normal voice) and an attempted imitation of Chinese-accented English. The actors were specifically asked to perform a Mandarin-Chinese accent. Although this request is obviously quite vague, it is typical of the level at which specific accents are discussed in materials for dialect courses and actor preparation for dialect performance (e.g. Blumenfeld 2002). Three of the actors – Emma, Matt, and Sango – had been students of the same acting coach at the University of Michigan. For their preparation, these actors used materials like Blumenfeld's lists to identify the important sounds that needed to change and then modified their scripted materials by adding IPA transcriptions of the affected words or, in most cases, individual sounds, they wished to alter. As can be seen below, these actors did not arrive at the same set of features, but, for the most part, were quite consistent in their alternations. Susie prepared for the role in this project by listening to and imitating recordings of Chinese-accented speech; in particular, she listened to recordings from the George Mason University speech accent archive, which provides accurate geographic information for each accented voice. Finally, Leo prepared for this role by listening to and imitating other actors. Indeed, Leo's performance was at times reminiscent of Mickey Rooney's portrayal of I. Y. Yunioshi in the 1961 film "Breakfast at Tiffany's."

One clear limitation of this study is the use of scripted materials. The actors' performances of Chinese-accented English contain none of the elided copulas or determiners that are both characteristic of authentic Chinese-accented English and present in Lindemann's description of folk linguistic ideologies. The actors focused instead on manipulations of pitch, duration, rhythm, and specific segmental alternations – as will my analysis. Actor-specific subsections

in the results section will detail the suprasegmental and segmental alternations made by each actor; however, results are likely to be quite different given a more improvisational task.

My prediction is that actors need only control features of Chinese-accented English that are socially informative to English-native listeners. To the extent that folk linguistic descriptions of an accent represent the linkages listeners have between social category and phonetic detail, the majority of actor manipulations should reflect Lindemann's respondents' folk descriptions, which were fairly low on both the Detail and Accuracy continua.

## Results

The first folk prediction is that Chinese-accented English is spoken quickly. If this feature is important for a native-English percept of "Chinese," then all of the actors are unsuccessful at producing it. Sentence durations were, on average, much longer (1,550 ms) for the actors in the Chinese condition than in their normal speaking voices (1,066 ms). Not all actors manipulated speaking rate, however. The actors identified here as Sango and Leo showed manipulation of speaking rate, while the actor identified as Emma spoke much more slowly.

Another folk prediction is that Chinese-accented English will be overall higher in pitch than English that is not Chinese-accented. A related prediction, not specifically tested with these materials, is that Chinese-accented English will rise in pitch when the speaker is cursing. There are no taboo words in these recordings and there was no request for particularly emotional speech. However, if we take these two predictions together as an expectation of both higher general F<sub>0</sub> and more variable F<sub>0</sub>, we can evaluate them in light of the collected materials. Surprisingly, there was no average or actor-specific manipulation of pitch in the creation of a Chinese-accented guise. Averaging mean F<sub>0</sub> for each sentence in the corpus, actors used a mean F<sub>0</sub> of 202 Hz in the Chinese Condition and 200 Hz in the normal speaking voice condition. The results are similar for the differences between minimum and maximum F<sub>0</sub> except for one actor, "Matt," who had a maximum pitch of nearly 400 Hz in his Chinese utterances, but only 210 Hz in his normal speaking voice.

## Segmental Alternations

Each actor performed at least a few consistent segmental alternations in their Chinese-accented guise. This is unsurprising given that these types of alternations are typically the primary focus in theatrical dialect training. The following subsections document each of the systematic segment-level

consonant replacements by each actor. One slight departure from normal convention is in the use of bracketing. In the following alternations, the slashes normally used for phonemic representations are used to present the actor's own normal productions, while the square brackets represent phonetic transcriptions of the segmental realization under the Chinese-accented condition.

*Susie*

/ð/	→	[d]			
/θ/	→	[t <sup>h</sup> ]			
/t/	→	{ [t <sup>h</sup> ] [t]	/	V__V	e.g. water
			/	__#	e.g. feet
/ɹ/	→	∅	/	V__	only in her and bird

Susie's aspiration of intervocalic and word-final [t] is correct insofar as it accurately reflects genuine difficulties L1 Mandarin speakers have with English (as described in the section entitled "Chinese-accented English," above). Susie's alternation of the interdental fricatives with their alveolar stop counterparts accurately recapitulates authentic Chinese-accented speech to the extent that L1 Mandarin speakers have difficulty with these fricatives. However, the more common alternation for authentic Chinese-accented speech is between the interdental and alveolar fricatives. Alternation with the stops does occur (indeed, it occurs in many varieties of foreign-accented English), but is more typical of, for example, Cantonese speech than Mandarin.

*Emma*

/ð/	→	[z]			
/l/	→	[w]			
/t/	→	[t <sup>h</sup> ]	/	V__V	e.g. water
/C#/	→	[Cə]		(optionally)	
/CC/	→	[CəC]			

Emma alternates the voiced interdental fricative with its alveolar counterpart in much the way that native L1 Mandarin speakers do. Her performance also faithfully recapitulates alternation between the alveolar lateral and the labiovelar approximant found in authentic speech, but she does this globally for all occurrences of /l/, whereas native L1 Mandarin speakers typically have this alternation only in word-final position. Finally, her aspiration of intervocalic [t] is correct in both alternation and environment. One striking way in which Emma's performance diverged from authentic Chinese-accented English was the frequent insertion of epenthetic schwa both as a means of cluster simplification and word-finally. This feature is not typical of L1 Mandarin-accented English, which simplifies by consonant deletion, but, like Susie's alternation of

the interdental fricatives with alveolar stops, is characteristic of other foreign-accented English varieties (e.g. Japanese, Cantonese, and Korean).

*Matt*

/l/	→	[w]		
/ɹ/	→	[w]		only in are
/t/	→	[tʰ]	/	V____V e.g. water

Matt, the only actor to manipulate mean pitch to any recognizable extent, employed the fewest segmental alternations of any actor. He accurately aspirated [t] inter-vocally. Like Emma, Matt alternated the alveolar lateral approximant with the labio-velar approximant, but, again, more uniformly than this is done in authentically Chinese-accented speech. Interestingly, Matt alternated the alveolar central approximant with the labio-velar approximant in the lexical item 'are'. This may reflect Matt's awareness of native difficulty with this approximant post-vocally, although he did not make this substitution more generally, nor did he delete post-vocally.

*Sango*

/ð/	→	[d]		(optionally)
/θ/	→	[s]	or	
		[tʰ]		
/s/	→	[s]	/	[+voice]____# e.g. falls, days, leaves, trees
/C#/	→	[Cʰ]	/	____# e.g. feet, fast, week
/ɹ/	→	∅	/	V____ e.g. dessert, her, sport, shirt, etc.
/dʒ/	→	[ʒ]	/	V____V e.g. pigeon

Sango's alternations were the most variable of any actor recorded for this study. Like other actors, Sango alternated the voiced interdental fricative with its alveolar stop counterpart. Like Leo, Sango consistently voiced the alveolar fricative word-finally – a pattern inconsistent with authentic Chinese-accented English and also, perhaps, incompatible with the authentic accent feature of word final devoicing of obstruents. However, several of her alternations were strikingly authentic. In particular, Sango was the only actor to lenite the voiced post-alveolar affricate to its fricative counterpart. She was also the only actor to delete /ɹ/ post-vocally both in consonant clusters and when it is alone in coda position in her unaccented voice. Aspects of her other alternations touch on authentic features of Chinese-accented English, but the distributions are either too narrow or too wide. The voiceless interdental fricative, for example, was often realized as an authentic voiceless alveolar fricative and, somewhat less often, as an aspirated voiceless alveolar stop. Authentic aspiration of inter-vocalic [t] was represented in Sango's performance, to a certain extent, by general word-final aspiration of voiceless stops.

*Leo*

/ð/	→	[d]				
/θ/	→	[t <sup>h</sup> ]				
/s/	→	[s]	/	[+voice]_____#		e.g. days, leaves, sleeves, trees
/t/	→	{	[t <sup>h</sup> ]	/	V_____V	e.g. water
			∅	/	C_____#	e.g. wrist, fast
/d/	→	[t]	/	_____#		e.g. head
/ɹ/	→	∅	/	V_____		e.g. her, water, sport

Leo's performance, while not intended in any way to be comedic or mocking, is clearly a blend of stereotypical Japanese- and Cantonese-accented English, not entirely dissimilar to the broad, comedic variety performed by actor Mickey Rooney in the film "Breakfast at Tiffany's" as "Japanese." Like several other actors, Leo stops the interdental fricatives and, like Sango, consistently voiced the alveolar fricative word-finally. His /t/ was either aspirated inter-vocally or deleted word-finally to simplify consonant clusters. This last feature is consistent with authentic Chinese-accented English, as were Leo's use of devoicing for word-final alveolar stops and post-vocalic deletion of the central approximant.

### Discussion

The prediction was that actors would perform primarily the highly salient, stereotypical features of a variety. Instead, the actors demonstrated surprising availability and control of a range of features which, while lacking significant overlap among them, shared many features in common, either with reported features of authentic L1 Mandarin-accented English or with aspects of those authentic alternations. Only one actor, Leo, imitated an accent clearly divergent from the requested Mandarin goal, and this accent was the result of studying other actors' performances of "Chinese." However, all actors produced imitated features suggestive of influence from Japanese, Cantonese, or Korean – suggesting a pan-ethnic Asian, rather than particularly Mandarin Chinese, set of social category representations. Perhaps, then, the level of awareness required to imitate a variety of foreign-accented English need only be consistent with the level of awareness one's audience is likely to have had during previous experience of that variety. Under normal circumstances, whether experiencing authentic Chinese-accented English or performed Asian/Chinese-accented English, it is unlikely that listeners are consciously aware of the precise national identity of the speaker. Actors not only reflect and appeal to these pan-ethnic social category representations, but create and reinforce them at the same time.

Neither the detailed, nor the general, folk predictions are well represented in these imitations. But one must wonder why actors might perform features of an



accent that are not required to invoke an accented percept for less experienced listeners. One possible answer is that, in fact, these listeners possess more detailed representations of Chinese-accented English than a map-labeling task can reveal. This interpretation is certainly supported by the results of Experiment 1, in which less experienced listeners performed better than chance when identifying authentically Chinese-accented English.

The results reported here suggest that less experienced listeners have a set of surprisingly detailed phonetic expectations linked to a pan-Asian social category. This knowledge is available for use during speech perception. Listeners may acquire this awareness of at least highly salient non-native phonological features through experience with imitated varieties, suggesting an important role for imitation and stereotype in listeners' use of the complex and informative patterns of variation available in the speech stream. Finally, this experiment offers further support for previous findings that listeners arrive at expectations of non-native features by drawing phonological analogies across social boundaries (e.g. Lindemann 2003).

### Conclusions

Speech perception is the use of linguistic knowledge to impose structure upon sensory data. There is abundant evidence showing that this linguistic knowledge is richly detailed. The present study is part of another growing body of research suggesting that what we think of as "linguistic" must be expanded to include knowledge of both phonetically cued social information and social categories – directly analogous to the well-established concepts of phonetically cued segmental or sublexical information and lexical categories.

When listeners are asked to identify an accented voice as "authentically Chinese" or when an actor seeks to create a "Chinese" percept in the minds of a native English-speaking audience, they attend to or manipulate phonetic cues that are simultaneously linguistic and socially meaningful. The single speech signal carries both meanings. Particular phonetic cues within that signal activate both linguistic (sounds and words) and social (speaker attributes, social category) representations. This finding is consistent with a growing body of work in speech perception (Beddor *et al.* 2013; Sumner and Kataoka 2013), psycholinguistics (Creel and Bregman 2011), and Sociolinguistics (Szakay *et al.* 2012), in which variation in speech – even quite dramatic variation from canonical forms – is reimagined as a source of information rather than a source of noise. Phonetic cues and social category knowledge interact to enhance the multiplex perception of referential and social indices.

In the authenticity detection task, more experienced listeners were better able to discriminate between an authentic and inauthentic variety. Less experienced listeners were more drawn to the imitated variety and, in the production

task, we learned that these imitated varieties index a simplified, pan-Asian social category that appears to be linked not only to expectations consistent with authentically Chinese-accented voices, but also Japanese and Korean-accented voices. This disparity highlights that, for exemplar models in which episodic memory is linked to social category information such as Johnson (e.g. 2006), awareness is required at some point during perception to link social category representations to stored exemplars. This is not meant to imply that listeners must have conscious awareness of the relationship between speaker attributes and particular phonetic cues to form a linkage between them; indeed, there is ample evidence that listeners are sensitive to phonetically cued social information below the level of conscious awareness (Koops *et al.* 2008; Nycz, this volume). Nor, I believe, does this claim contradict Squires's position (this volume) on perceiving versus noticing. Instead, the implication is that, under Johnson's model, it is impossible to form a linkage between a voice and a social category without having some awareness that the social category is salient. The phonetic variation can be learned without awareness and social cues can be learned without awareness, but the linkage from exemplar lexicon to higher level social category must require some awareness that the social category is available for linkage – even if that awareness lacks detail and accuracy. One take-away empirical prediction from the present chapter is that even thousands of hours of experience with Chinese-accented speech should not help a listener identify a voice as authentically Chinese without some awareness that the variety being experienced was in fact, or perhaps in stereotype, “Chinese.” One must not only attend to the signal, but one must attend to it as a representative signal of a particular social category.

Turning to the production study, only one actor imitated, in a limited way, the word-final obstruent devoicing which is a salient feature of authentically Chinese-accented English. This result, taken with the above, suggests that Labov's model of awareness accurately predicts even trained actors' use of the features of an imitated variant. Models of perception and representation such as Johnson (2006) or Sumner and Samuel (2009), in which accented productions or representations of accented variants are subsumed under a more general, standard representation, need to take into account Labov's model of awareness and the predictions it makes about access to these variants by speakers. The mixture of stereotypically Asian features produced by the actors in the imitated accent condition suggests that actors, and the listeners they hope to entertain, possess coarser-grained social categories of national identity than “Mandarin” versus “Cantonese” or even “Chinese” versus “Japanese.” These results suggest that listeners' conceptualization of this social category is much broader and all-inclusive.

The results reported here suggest that less experienced listeners have a set of surprisingly detailed, if somewhat pan-Asian, expectations available for use

during speech perception. These listeners may actually acquire this awareness of at least highly salient non-native phonological features through experience with imitated varieties; suggesting an important role for imitation and stereotype in listeners' use of the complex and informative patterns of variation available in the speech stream. Experienced listeners, on the other hand, as demonstrated by the results of Experiment 1, are much more accurate in identifying an authentic Chinese voice as "Chinese" and so are clearly drawing on a more fine-grained social category representation. The behavior of this experienced group of listeners is entirely consistent with exemplar models that include a linkage between knowledge of the speech signal and knowledge of speaker attributes. An interesting question not investigated in this study concerns the potential interplay of top-down stereotypical expectations and detailed bottom-up phonetic experience in the perception of listeners with extensive experience.

Finally, the ultimate goal of the chapter as a whole was to explore the feasibility of assessing and quantifying listeners' accumulated linguistic and social experiences for modeling experience in laboratory research. The authenticity detection task offers a means of assessing listeners' experience and, at least for Chinese-accented English, correlates well with listeners' self-reported experience labels. Since simply asking participants to self-report is easier and faster than administering an additional task, this task is of dubious utility for the investigation of phenomena which participants might be aware of and able to report. Furthermore, the results reported here highlight the fact that it can be difficult to tease apart quantity of experience with a target variety, in terms of raw frequency of exposure, from quality of experience, in terms of the accuracy of linkages from fine phonetic detail to social category and the structure or complexity of the social categories themselves. While the present task shows some promise, there is much room for improvement.

## REFERENCES

- Agha, A. (2003). The social life of cultural value. *Language and Communication* 23(3/4):231–73.
- Baayen, R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press.
- Baayen, R. H., Piepenbrock, R., and Rijn, H. V. (1993). *The CELEX Lexical Data Base on CD-ROM*. Philadelphia, PA: Linguistic Data Consortium.
- Babel, A. M. (2010). Contact and contrast in Valley Spanish. PhD thesis. Ann Arbor, MI: University of Michigan.
- Balota, D., Yap, M., Hutchison, K., Cortese, M., Kessler, B., Loftis, B., Neely, J., Nelson, D., Simpson, G., and Treiman, R. (2007). The English lexicon project. *Behavior Research Methods* 39(3):445–59.
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3):255–78.

- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1):1–48.
- Beddor, P., McGowan, K., Boland, J., Coetzee, A., and Brasher, A. (2013). The time course of perception of coarticulation. *Journal of the Acoustical Society of America* 133(4):2350–66.
- Beddor, P. S., Harnsberger, J., and Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: acoustic structures and their perceptual correlates. *Journal of Phonetics* 30(4):591–627.
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model, in J. C. Goodman and H. C. Nusbaum (eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, pp. 167–224. Cambridge, MA: MIT Press.
- Blumenfeld, R. (2002). *Accents: A Manual for Actors*. New York: Proscenium Publishers Incorporated.
- Bradlow, A. and Alexander, J. (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):2339–49.
- Brunner, E. G. (2010). Imitation, awareness; and folk linguistic artifacts. PhD thesis. Houston, TX: Rice University.
- Bucholtz, M. and Hall, K. (2010). Locating identity in language, in C. Llamas and D. Watt (eds.), *Language and Identities*, pp. 18–28. Edinburgh University Press.
- Campbell-Kibler, K. (2009). The nature of sociolinguistic perception. *Language Variation and Change* 21(1):135–56.
- Creel, S. and Bregman, M. (2011). How talker identity relates to language processing. *Language and Linguistics Compass* 5(5):190–204.
- Dahan, D., Drucker, S. J., and Scarborough, R. A. (2008). Talker adaptation in speech perception: adjusting the signal or the representations? *Cognition* 108:710–18.
- Docherty, G. J. and Foulkes, P. (2014). An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua* 142:42–56.
- Drager, K. (2010). Sociophonetic variation in speech perception. *Language and Linguistics Compass* 4(7):473–80.
- Eckert, P. (2008). Variation and the indexical field. *Journal of sociolinguistics* 12(4):453–76.
- Espiritu, Y. (1992). *Asian American Panethnicity: Bridging Institutions and Identities*. Philadelphia, PA: Temple University Press.
- Foulkes, P. and Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics* 34:409–38.
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6:110–25.
- Gick, B. and Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature* 462:502–4.
- Hay, J., Warren, P., and Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34:458–84.
- Irvine, J. (1989). When talk isn't cheap: language and political economy. *American Ethnologist* 16(2):248–67.
- Iwamoto, D. K. and Liu, W. M. (2010). The impact of racial identity, ethnic identity, Asian values, and race-related stress on Asian Americans and Asian international

- college students psychological well-being. *Journal of Counseling Psychology* 57(1):79.
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model, in K. Johnson and J. W. Mullennix (eds.), *Talker Variability in Speech Processing*, pp. 145–65. San Diego, CA: Academic Press.
- (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34:485–99.
- Johnstone, B. and Kiesling, S. F. (2008). Indexicality and experience: Exploring the meanings of /aw/-monophthongization in pittsburgh1. *Journal of Sociolinguistics* 12(1):5–33.
- Koops, C., Gentry, E., and Pantos, A. (2008). The effect of perceived speaker age on the perception of pin and pen vowels in Houston, Texas. University of Pennsylvania Working Papers in Linguistics: Selected papers from NWAV 36, 14(2):91–101.
- Kucera, H. and Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence, RI: Brown University Press.
- Kuznetsova, A., Brockhoff, P., and Christensen, R. (2013). lmerTest: tests for random and fixed effects for linear mixed effect models. R package version 2.0–11.
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.
- Liberman, A. M., Harris, K. S., and Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54:358–68.
- Lindemann, S. (2003). Koreans, Chinese or Indians? Attitudes and ideologies about non-native English speakers in the United States. *Journal of Sociolinguistics* 7(3):348–64.
- (2005). Who speaks “broken English”? US undergraduates’ perceptions of non-native English. *International Journal of Applied Linguistics* 15(2):187–212.
- Lisker, L. (1986). “Voicing” in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29:3–11.
- Mack, S. and 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.
- Mann, V. A. and Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America* 69:548–58.
- McGowan, K. B. (2015). Social expectation improves speech perception in noise. *Language and Speech*. 58(4):502–21.
- McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264:746–8.
- Milroy, L. and McClenaghan, P. (1977). Stereotyped reactions to four educated accents in ulster. *Belfast Working Papers in Language and Linguistics* 2:1–11.
- 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* 1:157–77.
- Neuhausser, S. and Simpson, A. P. (2007). Imitated or authentic? listeners’ judgements of foreign accents, in *ICPhS XVI*. Saarbrücken: International Congress of Phonetic Sciences.

- Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18(1):62–85.
- Pierrehumbert, J. B. (2002). Probabilistic phonology: Discrimination and robustness, in R. Bod, J. Hay, and S. Jannedy (eds.), *Probability Theory in Linguistics*. Cambridge, MA: MIT Press.
- Preston, D. R. (1996). Whaddayaknow?: The modes of folk linguistic awareness. *Language Awareness* 5(1):40–74.
- Qiong, H. X. (2004). Why China English should stand alongside British, American, and the other. *English Today* 20(02):26–33.
- R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07–0.
- Rubin, D. L. (1992). Non-language factors affecting undergraduates' judgments of non-native English-speaking teaching assistants. *Research in Higher Education* 33(4):511–31.
- Schulman, R. (1983). Vowel categorization by the bilingual listener. PERILUS Working Papers, III:81–99.
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication* 23(3–4):193–229.
- Sumner, M. and Kataoka, R. (2013). Effects of phonetically-cued talker variation on semantic encoding. *Journal of the Acoustical Society of America* 134(6): EL485–91.
- Sumner, M., Kim, S. K., King, E., and 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. and 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., and 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):138–46.
- Tang Tang, C. C. and van Heuven V. J. (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua* 119(5):709–32.
- Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., and 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):510–40.
- Wells, J. (1982). *Accents of English*. Cambridge University Press.
- Wong, P., Lai, C. F., Nagasawa, R., and Lin, T. (1998). Asian Americans as a model minority: Self-perceptions and perceptions by other racial groups. *Sociological Perspectives* 41(1):95–118.