IMITATION IN SPEECH

Molly Babel
Department of Linguistics
University of British Columbia
Vancouver, British Columbia V6T 1Z4, Canada

> *"Social preferences and liking modulate the process of spontaneous phonetic imitation."*

Why do we sound the way we do? As people learn to speak, they acquire the language and dialect spoken around them. Sentence structure, word choice, and pronunciation are all determined by the patterns used in the ambient language to which we are exposed. Having grown up in Minnesota, I did not learn how to speak with a British accent, but with a Minnesotan one. As a language-learning child, the language input I received determined the general shape of my language output. This article focuses on the spontaneous or natural imitation of speech acoustics. In this article I use the terms imitation, convergence, and accommodation interchangeably. I use all of these terms to describe the unintentional process by which exposure to a speech stimulus causes an observer to display characteristics of the stimulus in their own productions.

This phenomenon of imitating the input is not limited to the acoustic signal that we use to transmit language. Let's begin with a straightforward example from the literature on syntactic priming and word order of how recent linguistic exposure modifies our subsequent speech behavior. Imagine a picture of a man holding a cake and facing a woman. The orientation of the image suggests the man intends to pass the cake to the woman. Participants who have been exposed to the sentence *The boy gave the toy to the teacher* prior to viewing this image are more likely to describe the cake picture as *The man gave the cake to the woman* as opposed to *The man gave the woman the cake*. The second description is a completely grammatical utterance that accurately conveys what is going on in the image, but having been previously exposed to the construction *give X to Y* biases the future use of that construction over *give Y X* (Bock, 1986). Bock's seminal finding reveals quite convincingly that what we say is highly influenced by what we have just heard.

The notion that children learn the speech variety to which they are exposed seems intuitive, but the situation becomes a little more complicated when we shift our attention to the acquisition patterns of adults. What happens when adults who have already acquired a particular speech variety move to a new dialect area? As a young adult, I moved to California and, with time, my speech lost many of its original Minnesotan features. To native Californian ears, I might never have sounded truly Californian, but I eventually sounded much less Minnesotan. Several recent studies have documented this personal anecdote on a larger scale (Evans and Iverson, 2006; Munro *et al.*, 1999). Interestingly, the Minnesotan features of my speech return when I am interacting with old friends and family who have retained our native dialect. This indicates that the acquisition of a new dialect or the adoption of new speech features serve as an update to and expansion upon my linguistic system, as opposed to wholly replacing a previous system. The fact that we do imitate the ambient language tells us that our linguistic categories are malleable and easily influenced by new information.

Inherent variability in speech production

When one considers how speech is produced, it becomes apparent that physiological and anatomical variation across talkers will inevitably be reflected in the spectral characteristics of speech sounds. Let's start with the sex of the talker. Men's voices typically pattern together in having lower pitch and lower resonant frequencies than women's voices. These differences are due in part to sexual dimorphism: men's vocal tracts and vocal folds are generally larger than women's. Age is another key factor in cross-talker variability. Aging is accompanied by various physiological changes. For example, the extrinsic muscles that support the larynx become slack with age, and the mucosal tissue covering the vocal folds loses its elasticity. These changes lead to alterations in voice quality, along with lower pitch and lower resonant frequencies for a talker's voice. These factors highlight the fact that some of the variation in speech is due to anatomical and physiological age- and sex-based variation.

That being said, the extent to which physiological factors determine speech characteristics is frequently oversold. A closer inspection of large datasets reveals that physiological variation does not account for all of the observed differences between groups divided by gender or age. While it is true that classic studies describing vocalic resonant frequencies of men and women show that women produce higher resonant frequencies than men (Peterson and Barney, 1952), a recent examination of gender differences across languages illustrated that gender-based differences in vowel production vary across languages, even when population height is controlled (Johnson, 2006). Such a finding indicates that some of the gender-based differences in resonant frequencies are the result of learned social norms. In addition, despite the fact that many significant anatomical differences do not emerge between males and females until puberty, children acquire gender-specific speech patterns starting in toddlerhood (Sachs *et al.*, 1973; Perry *et al.*, 2001). This suggests a strong socio-cultural component to language production.

Armed with this information, we can hedge a response to the question of why we sound the way we do by acknowledging that a portion of one's speech acoustics is determined by the size and shape of the vocal tract. There is clearly more at issue, however. As mentioned above, we acquire the speech

SPEECH COMMUNICATION 101

You say tomato, I say tomato

The types of changes to which I refer in this paper involve sometimes subtle, sometimes not-so-subtle, changes in the pronunciation of particular sounds. It is important to familiarize ourselves with the ways in which linguists and speech scientists talk about speech sounds. Speech production boils down to a manipulation of the airstream. For example, say the word *tomato*; do this with your hand in front of your mouth and you will have a tactile impression of this airstream manipulation, in addition to the auditory one. Figure 1 presents a spectrogram and waveform from two speakers' pronunciations of *tomato*. My production is on the left, and a male's production of this word is on the right. We return to some key differences in female and male productions later.

Tomato, like all words, is made up of a series of speech sounds. To produce the word *tomato*, your tongue tip moves up and makes contact in the region behind your front teeth, sometimes making contact with your teeth themselves, to make the /t/. If you put your hand in front of your mouth, you will feel a rather strong puff of air as you release the /t/; this is called *aspiration*. From the /t/, your mouth changes its configuration seamlessly as it moves towards a more neutral configuration for the initial vowel, which is a shortened, indistinct schwa-like vowel. Then, to produce an /m/, you close your lips and open the passageway to your nasal cavity, allowing the air to flow through your nasal cavity and sinuses on the way to the open atmosphere. Following the /m/, the vowel you produce will vary considerably depending on the variety of English you speak. Most speakers from North America will produce the vowel sound which also occurs in *bake* and *cake*: /e͡ɪ/. If you speak a variety of British English, you will likely produce this sound as an /ɑ/, which is more similar, although not identical, to the vowel that North American English speakers use in the word *father*. Next we come to a sound that is like /t/, but which is produced much more quickly and with a different movement trajectory in natural speech; this is called a *flap*: /ɾ/. Finally, there is an /o/ sound, which involves rounding your lips a little bit. Note that this last vowel sound is produced and sounds quite different from the first vowel in the word, despite the fact they are both spelled with an "o." It is rather amusing how long it takes to describe *how* to produce a word, compared to how long it takes to simply say the word. Our mouths do some impressive articulatory gymnastics at incredible speed in speech production, and we do not even give it much, if any, thought.

These articulatory movements modify the airstream, making constrictions in the oral cavity to varying degrees. In making a /t/, there is a complete constriction to the point where air is trapped inside the oral cavity. To produce a vowel sound, the oral cavity is left relatively unconstricted, but the oral cavity and tongue are shaped in particular ways. These constrictions and configurations result in particular acoustic consequences when the air column in your vocal tract is excited. You can move your articulators to your heart's content, but without the excitation of the air column in your vocal tract, no sound will be emitted.

A crash course on the acoustics of speech production

The acoustics of the speech signal are determined by two main factors: the sound source and the filter through which that sound passes. The role of the filter is to modify the spectral shape of what was produced by the sound source. In the production of voiced sounds with a relatively open vocal tract—sounds like vowels, /r/, and /l/—the

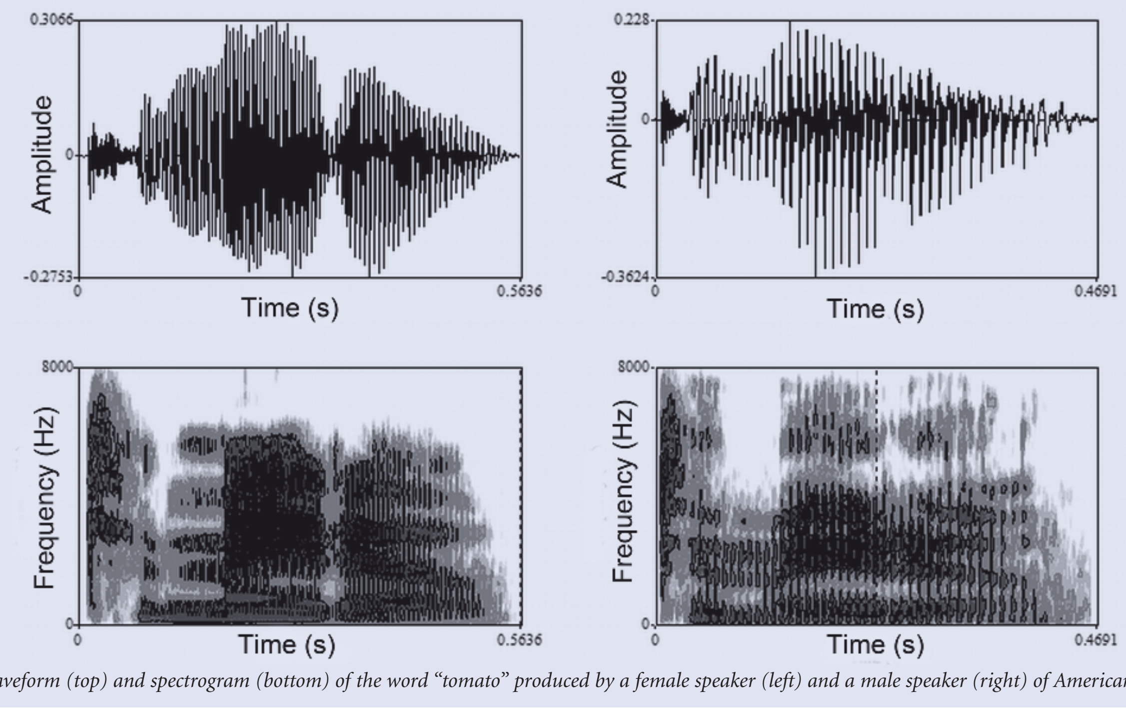

Fig. 1. A waveform (top) and spectrogram (bottom) of the word "tomato" produced by a female speaker (left) and a male speaker (right) of American English.

vibration of the vocal folds serves as the sound source. The supraglottal cavity, which is the part of the vocal tract that extends above the vocal folds, is the filter which shapes the sound generated at the vocal folds. Differences in the size and shape of the vocal folds contribute to inter-speaker differences in pitch and voice quality. Larger vocal folds with more mass will vibrate more slowly, producing a voice that is lower in pitch than that of a talker with smaller vocal folds. The oscillation of the vocal folds provides the fundamental frequency, which listeners perceive as the pitch of the voice, and harmonics, which occur at multiples of the fundamental frequency. “Voice quality” refers to the variations in the sound of a talker’s voice, ranging, for example, from breathy to modal to creaky. It is determined largely by the relative speed and duration of vocal fold closure in the course of vibration, and by whether full vocal-fold closure is achieved. (The vibration of the vocal folds is a complex process and we are glossing over many details here.) The size and morphology of the supraglottal cavity determines the resonant properties of the filter. The larger this oral cavity, the lower the resonant frequencies produced by the vocal tract. Manipulating the configuration of the vocal tract modifies the resonant frequencies it produces. These resonant frequencies provide valuable information about a speaker’s vowels within the acous-tic-phonetic space of human vocalizations.

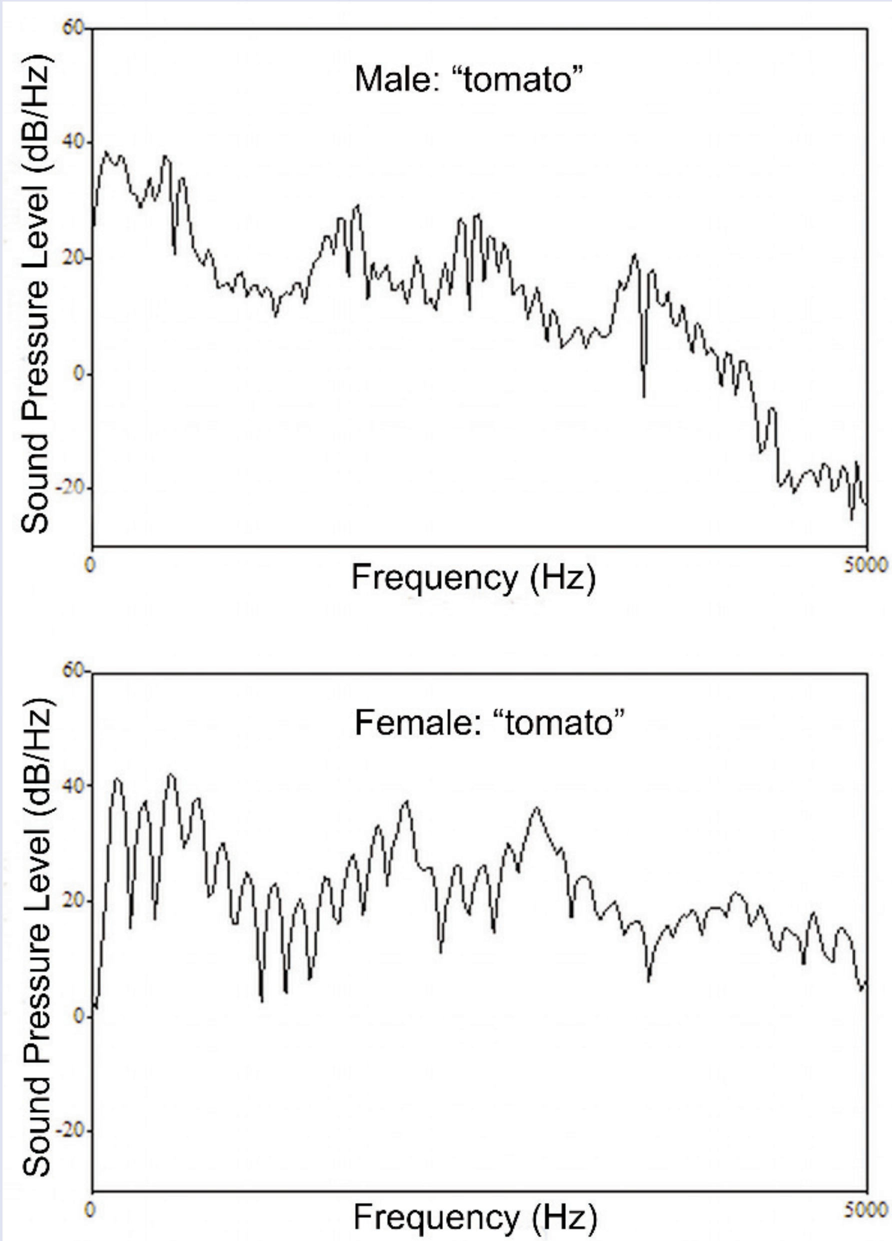

Fig. 2. Power spectra from the first third of the final vowel in the word tomato from a male (top) and a female (bottom) speaker.

Power spectra made from the first third of the final vowel in *tomato* from a male (top) and female (bottom) speaker are shown in Fig. 2. There are several differences between the female and male productions. Let’s first turn our attention to the lowest frequency component of the spectra. This is the fundamental frequency. The fundamental frequency is lower in the male’s voice than in that of the female. The components going up in frequency are the harmonics. The fundamental and the harmonics are produced by the vibrating vocal folds. Note also that the high amplitude resonance peaks in the female-produced spectrum are generally higher in frequency than the high amplitude resonance peaks in the male-produced spectrum. One intuitively expects to find these predictable differences in speech produced by males and females, due to the relative differences of their vocal tract sizes. Differences in filter shape and source vibrations, however, would arise even when comparing spectra for two female speakers, or two productions from the same speaker. The number of potential acoustic parameters is limitless, and speakers do not have complete control over all of these parameters. Speech, simply, is immensely variable.

characteristics of those around us, so let's consider for a moment regional and dialect variation in speech production. Imagine modeling the pronunciation variants of the vowel /u/, as in the word *food* or *dude*, for which there is considerable variation in regional vernacular production. In Minnesota, this vowel is produced with the tongue in a very high and very back position; the lips are also considerably rounded. These articulatory movements cause this vowel to have a very low second resonant frequency. Now, picture to yourself a Californian surfer saying *dude*. In this stereotypical version, production is quite different from the Minnesotan *dude*. This type of /u/ pronunciation has an extremely high second resonant frequency. Minnesotans and Californians will not globally produce these vowels in exactly these caricatured ways; this increase in the second resonant frequency of /u/ is part of a sound-change-in-progress, leaving individuals in all communities at different stages of the change.

Talk show hosts imitating their guests

The acquisition of regional variation can be simply considered part of what we acquire based on what we hear around us. However, language use does not only vary according to region, and we also find systematic variation based around other macro-sociological categories like class and ethnicity. Our use of language does not reflect a monochromatic mirroring of what we acquired as children, but rather a flexible matching process of sorts that is largely influenced by who we are speaking with or what we are talking about. Let's take as an example the speech patterns of two celebrities. Consider first the pronunciation patterns of Ms. Oprah Winfrey from the *Oprah Winfrey Show*, which were analyzed in Hay *et al.* (1999). Under analysis was the degree of monophthongization of /ɑ͡ɪ/ in Ms. Winfrey's speech; that is, did she pronounce a word like *time* as /t^{h}ɑ͡ɪm/ or /t^{h}am/? Examples of these two variants are shown in Fig. 3. The primary acoustic differences between these two variants relate to the first and second resonant frequencies of the vocal tract. Note, for example, the dynamic trajectories in /t^{h}ɑ͡ɪm/ on the left compared to the more stable resonant frequencies in /t^{h}am/ on the right. Given Ms. Winfrey's linguistic biography, we might predict her speech would make use of both variants. Ms. Winfrey grew up in rural Mississippi, a region where /ɑ͡ɪ/ monopthongization is common. This is a typical feature of African American English-speaking speech communities as well. Given the racial division in the Southern US during her childhood, we can infer that Ms. Winfrey grew up in a speech community where this monophthongization was common. As an adult, Ms. Winfrey lives and interacts in speech communities where this sort of monophthongization is not common, and where most talkers use a diphthongal pronunciation.

Hay and colleagues demonstrated that the way Ms. Winfrey pronounced this word varied as a function of both how frequently the word was used and the racial identity of her upcoming guest. Words were defined as high frequency if used five or more times in the corpus under study, and low frequency if used fewer than five times. The researchers found words used more frequently were more likely be produced with the monophthong /t^{h}am/: 30% of the frequent words were monophongized, as opposed to only 14% of the infrequent words. The racial identity of the upcoming guest largely influenced Ms. Winfrey's pronunciation of this vowel as well. Ms. Winfrey was three times more likely to use a monophthongal pronunciation when she was introducing or discussing an upcoming African American guest on her television program than when she was talking about a non-African American guest. We can interpret her behavior as an accommodation process where she uses a particular variable based on the predicted pronunciation patterns of her guests. This is a process of a sort of global speech style imitation or accommodation. The fact that lexical frequency influences the pattern is also important. The pronunciation variability is not wholly determined by social and interpersonal context; language internal factors, such as lexical frequency, also affect how sounds are produced.

A second celebrity example of acoustic imitation takes us to another talk show: *Larry King Live*. On the *Larry King Live* show, Mr. King interviewed a range of guests, including celebrities, politicians, and others. Gregory and Webster

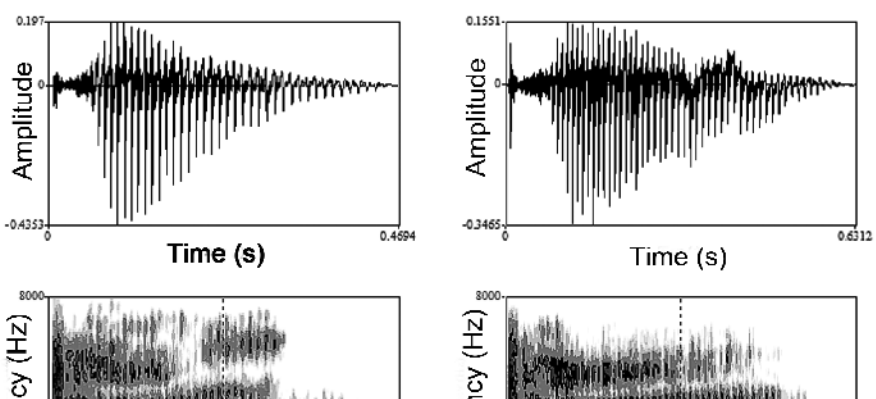

Fig. 3. Pronunciation variants of the word time. The variant on the left is the diphthongal /t^{h}ɑ͡ɪm/ variant and the token on the right is the monophthongal /t^{h}am/ variant.

(1996) took speech samples from Mr. King and twenty-five of his guests, measuring the long-term average spectra (LTAS) in a low frequency band-pass filter from each speech sample. Using these LTAS measures, correlation coefficients of the actual conversations were compared to those of pseudo-conversation pairings. Actual conversations had significantly higher correlation coefficients, suggesting that interacting talkers accommodated their spectral patterns. The researchers used a measure of LTAS variability and established that Mr. King's interviewees fell into a dominating, low deference group and a high deference group. Mr. King's LTAS measures indicated he took a deferent stance toward the dominating group and a more dominating position in interviews with those in the high deference group. Influential politicians of the time, like former President George H. W. Bush and former President Bill Clinton, fell into the dominating group, triggering a more deferring response from Mr. King. On the other hand, former Vice President Dan Quayle was a member of the high deference group who accommodated more to the speech patterns of Mr. King. (These interviews were taken from broadcasts from April 1992 through July 1993; it may help to keep the social and political context of that era in mind.) Undergraduate students completed surveys to rate the social status of the interviewees and to evaluate how perceived social status affected Mr. King's accommodative behavior toward the interviewees. These subjective student-elicited measures echoed Mr. King's patterns of phonetic accommodation: for example, former President Clinton was at the top of the social status ranking, while former Vice President Quayle was at the bottom. Simply, the researchers found that Mr. King accommodated to the speech patterns of his guests who were of higher social status, while the lower status guests accommodated low frequency spectral characteristics of their speech toward those of Mr. King.

Phonetic imitation in the speech laboratory

Talk show hosts are, of course, not the only ones who imitate and accommodate to interlocutors during spoken language interaction. In recent years, phonetic imitation has been a hot topic in laboratory-based studies of speech. These studies often take one of two forms: a task guised as a sort of guided spontaneous conversation or an auditory naming task. Let's discuss these in turn, starting with the guided spontaneous conversations. These are often map tasks or spot-the-difference tasks, typically involving two participants. They are guided in the sense that they are centered around cooperative activities dictated by the experimenter, but are spontaneous in that the detail of the conversation is freely determined by the natural interaction. In one recent and influential study on phonetic convergence, Pardo (2006) examined phonetic convergence in same-gender dyads involved in jointly completing a map task, where one member of the dyad was the giver of map directions and one was the receiver whose task it was to navigate the described path. Dyads were found to have converged on 62% of the experiment trials. Female dyads were found to converge toward the speaker who was receiving instructions, whereas male dyads patterned oppositely; they converged toward the speech of the talker giving instructions. Pardo concluded that particular social factors dependent on the situational context of a conversation—factors such as gender and the power dynamic of a giving-receiving interaction—determine the direction of phonetic accommodation. Another recent study examined convergence between dyads completing a spot-the-difference task (Kim *et al.*, 2011). The conversational dyads in this study were pairs of native English speakers and native Korean speakers who either did or did not speak the same dialect, and dyads of native and non-native speakers of English. More accommodation was found in the same-dialect dyads than in the different-dialect or the cross-language pairs. Kim and colleagues concluded the process of convergence is facilitated when members of a dyad share a language background, indicating that convergence is easier when the target of the convergent behavior is within an individual's pre-existing phonetic repertoire.

The second design frequently employed in the literature is an auditory naming task. An auditory naming task consists of a listener hearing a model talker produce a word over headphones or loudspeakers, and the listener's task is to identify the word by saying it out loud: that is, to *name* the auditory object. While this method eliminates the natural social context for imitative behaviors to emerge, it offers a more controlled environment for speech researchers to query particular aspects of what might facilitate or inhibit the imitation process. Using this methodology, Goldinger (1998) established that less common words are imitated more than words that are used more frequently. This finding suggests that phonetic imitation may play a role in how we learn about our native languages. For example, the number of times you have heard the word *potato* uttered around you is likely many times more than the number of times you have heard *kohlrabi*. Exposure to variation in how *potato* is pronounced is unlikely to sway your production of the word: in a sense, you confidently know how to say *potato*. Hearing a slightly different pronunciation of the word *kohlrabi*, on the other hand, may cause you to cast your own pronunciation of the word into doubt. On some level, this is an inaccurate way to describe the role of lexical frequency in phonetic imitation; an auditory naming task is too fast-paced to allow for personal reflection on pronunciation insecurities. However, if you have amassed fewer experiences with a particular word, you have fewer memories about how to pronounce it, making a single new exposure to *kohlrabi* all the more prominent a perceptual experience.

Other research using an auditory naming paradigm has explored the interaction of social and phonetic factors in speech imitation. In my own work, I have examined how social biases and preferences moderate phonetic imitation. In an auditory naming task where New Zealanders were presented with an Australian model talker, implicit social biases in New Zealanders' positive or negative views about Australia were measured using an Implicit Association Task, a standard social psychology tool (Babel, 2010). A strong relationship was found: while overall New Zealanders imitated the Australian model, the more positive the New Zealanders' implicit social biases toward Australia, the more they imitated. In another study, I found that the more attractive female participants rated a male model talker, the more they imitated his vowels (Babel, 2012). Social preferences and liking thus modulate the process of spontaneous phonetic imitation. This suggests that the relationship between speech perception and production may be somewhat labile in nature. In a study of gender bias in imitation, Namy and colleagues found that women imitated more than men, but that women's imitative behaviors were focused on a particular male voice used in the experiment (Namy *et al.*, 2002). This indicates there was a particular aspect of this male model's voice which encouraged females to imitate it more than the other voices used in the task.

It is clear that social factors play a role in phonetic imitation. Language-specific internal factors are also involved in the process of imitating speech, just as lexical frequency played a role in Ms. Winfrey's variable pronunciation of /ɑ͡ɪ/ or /a/. Using a modified version of an auditory naming task, Nielsen (2011) presented listeners with a block of model productions that had been digitally modified such that the aspiratory puff of air that accompanies the /p^h/ sound in *pint* or *pound* was longer than it typically is in natural speech production. (Note that such a puff does not occur in /b/ initial words like *beer* or *baseball*.) Nielsen found exposure to these modified words caused participants to not only increase the duration of the aspiration of /p^h/, but also to generalize this increase in aspiration to /k^h/ initial words like *canoe* and *kite*. This indicates that imitation can be abstracted and generalized across one's linguistic system.

Measuring imitation

How do researchers determine whether imitation took place? The complexity of the speech signal makes the method of measuring imitation an important topic. There are two primary ways to gauge or measure phonetic imitation—acoustic or perceptual—each with its positive points and drawbacks. The choice between them is guided by the goals of the study. Let's say a researcher would simply like to demonstrate that the speech signal was imitated in some way. The most common way to accomplish this is to have naive listeners rate perceptual similarity using an AXB task, in which listeners are presented with three tokens in a trial and are asked to judge whether the A or B token is more similar to the X token. The X token is a speech sample from Talker 1, the model talker who was presumably imitated. The A and B tokens would be speech samples from Talker 2; one token would be a "baseline" sample recorded before exposure to or interaction with Talker 1, and the other token would be a sample recorded during or after exposure to or interaction with Talker 1. When listeners consistently choose the post-exposure token as more similar-sounding to the X production of the same word, there is evidence for phonetic imitation. There are two primary virtues to using an AXB similarity task to assess phonetic imitation: (1) it is a holistic measure that allows for imitation of any part of the acoustic signal to contribute to listener judgments of perceived similarity, and (2) if imitative behaviors serve as a seed to sound change, as has been argued (Garrett and Johnson, in press), then these behaviors must be perceptible to listeners.

Oddly enough, this issue of sound change is critical in choosing to measure imitation acoustically as well. It has been argued that phonetic imitation is the means by which sound changes spread through communities. Sound change tends to affect particular aspects of the speech signal. Let's return to the earlier example of the pronunciation of /u/. It was noted above that the pronunciation of this vowel in words like *dude* varies as a function of region. In addition to this regional variation, there is a sound change in progress with this vowel across most varieties of North American English. The sound change is not random: the second resonant frequency of this vowel is becoming higher due to talkers adopting a more fronted tongue position. Acquiring specific acoustic evidence about what was imitated has the potential to address the issue of whether phonetic imitation is the seed by which sound change is spread. In imitating words with /u/, is this increase in the second resonant frequency one of the acoustic features that listeners imitate? If imitation studies demonstrate that what is imitated corresponds to the acoustic phonetic details involved in sound change, then researchers have promising evidence for how sound changes might spread through speech communities. While specific acoustic measures of imitation offer valuable insight into *what* is imitated, perceptual measures of imitation attenuate a potential experimenter bias: in acoustic measures of imitation, the researcher must predict which aspects of the signal are worth measuring, whereas with a holistic perceptual approach, all features of the signal can be taken into consideration. Recent work comparing a single acoustic measure of imitation and perceptual measures of imitation demonstrated that even when both measures reveal significant effects of imitation, the acoustic and perceptual measures are not correlated (Babel and Bulatov, 2011). This finding underscores the fact while phonetic imitation may interact with individual acoustic-phonetic features on the macro-level in terms of the diffusion of sound change across communities, phonetic imitation on an individual level does not involve singular features. Furthermore, this result indicates that listeners naturally evaluate perceptual similarity from a more holistic perspective. Of course, more holistic acoustic measures are also possible. Recent work, for example, has used mel-frequency cepstral coefficients as a measure of phonetic imitation (Delvaux and Soquet, 2007).

Concluding remarks

The research on phonetic imitation allows several important conclusions about speech communication. First, it highlights an important feature about speech perception and speech production. For imitation to occur, listeners have to perceive a certain amount of subtle acoustic-phonetic detail in the speech signal. This underscores listener sensitivity to the details of the signal. From there, the listener-turned-talker must map the acoustic-phonetic detail onto their own subsequent speech productions. This observation from the imitation literature indicates a relationship between speech perception and production. Importantly, this relationship cannot be a one-to-one mapping because we find many cases where phonetic imitation does not occur (see Vallabha and Tuller (2003) for a clear example). Second, work on phonetic imitation brings us to an interesting conclusion with respect to social cognition and language. The data indicate that, at least in laboratory contexts where social meaning is comparatively void, the default behavior seems to be imitation. Outside of the laboratory, we can imagine that imitation would be crucial for creating and developing social cohesion. The fact that socio-cultural factors moderate even low-level laboratory-based speech behavior strongly suggests that speech production is never without social influence. Lastly, the syntactic analogue of phonetic imitation, known as syntactic alignment or priming (previewed in the introduction), alludes to the important observation that imitative behaviors are pervasive across the language system. This suggests that imitation may serve as a fundamental component in the process of language acquisition and language learning.**AT**

References

Babel, M. (**2012**). "Evidence for phonetic and social selectivity in spontaneous phonetic imitation," J. Phonetics, **40**, 177–189.

Babel, M. (**2010**). "Dialect divergence and convergence in New Zealand English," Language in Society **39**, 437–456.

Babel, M., and Bulatov, D. (**2011**). "The role of fundamental frequency in phonetic accommodation," to appear in Language and Speech, 1–18, doi:10.1177/0023830911417695.

Bock, K. (**1986**). "Syntactic persistence in language production," Cognitive Psychology **18**, 355–387.

Delvaux, V., and Soquet, A. (**2007**). "The influence of ambient speech on adult speech productions through unintentional imitation," Phonetica **64**,145–173.

Evans, B. G., and Iverson, P. (**2007**). "Plasticity in vowel perception and production: A study of accent change in young adults," J. Acoust. Soc. Am. **121**, 3814–3826.

Garrett, A. and Johnson, K. (**in press**). "Phonetic bias in sound change," in A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (The Oxford University Press, Oxford).

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychological Rev. **105**, 251–279.

Gregory, S.W., and Webster, S. (**1996**). "A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions," J. Personality and Social Psychology **70**, 1231–1240.

Hay, J., Jannedy, S., and Mendoza-Denton, N. (**1999**). "Oprah and /ay/: Lexical Frequency, Referee Design and Style," *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, CA.

Johnson, K. (**2006**). "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," J. Phonetics **43**, 485–499.

Kim, M., Horton, W. S., and Bradlow, A. R. (**2011**). "Phonetic convergence in spontaneous conversations as a function of interlocutor language distance," J. Laboratory Phonology **2**, 125–156.

Munro, M. J., Derwing, T. M., and Flege, J. E. (**1999**). "Canadians in Alabama: A perceptual study of dialect acquisition in adults," J. Phonetics **27**, 385–403.

Namy, L. L., Nygaard, L. C., and Sauerteig, D. (**2002**). "Gender differences in vocal accommodation: The role of perception," J. Lang. and Social Psychology **21**, 422–432.

Nielsen, K. (**2011**). "Specificity and abstractness of VOT imitation," J. Phonetics **39**, 132–142.

Pardo, J. S. (**2006**). "On phonetic convergence during conversational interaction," J. Acoust. Soc. Am. **119**, 2382–2393.

Perry, T. L., Ohde, R., and Ashmead, D. (**2001**). "The acoustic bases for gender identification from children's voices," J. Acoust. Soc. Am. **109**, 2988–2998.

Peterson, G., and Barney, H. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Sachs, J., Lieberman, P., and Erickson, D. (**1973**). "Anatomical and cultural determinants in male and female speech," in *Language Attitudes*, edited by, R. W. Shuy and R. W. Fasold, 74–83 (Georgetown University Press, Washington, DC).

Vallabha, G. K., and Tuller, B. (**2004**). "Perceptuomotor bias in the imitation of steady-state vowels," J. Acoust. Soc. Am. **116**, 1184–1197.

Molly Babel is an Assistant Professor in the Department of Linguistics at the University of British Columbia. She teaches courses in phonetics, laboratory phonology, and social factors in language use. Her research examines variation in the speech signal with particular attention to the relationship between speech perception and production and how social cognition modulates speech processing. She maintains multiple lines of research which investigate several sides of these issues including imitation and accommodation, audio-visual and cross-linguistic speech perception, and the effects of talker typicality in speech processing. Babel received her BA in Linguistics, Anthropology, and Spanish from the University of Minnesota, Twin Cities, and her MA and PhD degrees in Linguistics from the University of California, Berkeley.