Sensitivity to Foreign Accent

Kenneth J. de Jong

Address:
Department of Linguistics
Indiana University
1020 E. Kirkwood Avenue
859 Ballantine Hall
Bloomington, Indiana 47405
USA

Email:
kdejong@indiana.edu

Listeners are extraordinarily sensitive to acoustic detail in the speech of nonnative speakers.

Research on the acoustics of the speech of nonnative speakers shows how extraordinarily sensitive listeners are to the very subtle details of speech patterns. When we hear a person speaking, we very rapidly not only hear what the speaker is saying but also are cognizant of many other attributes of the speaker: whether male or female, emotional state, whether the speaker has a cold, and whether the speaker is a native member of our own speech community. This article reviews a variety of work on the perception of whether a person is a native speaker or not and points out the remarkable sensitivity that the listeners have to information about accent. From research in perceiving the native quality of speech, we might say that speech acoustics not only captures our attention as scholars but also captures the rapt attention of everyone.

The obvious reason for this attention is that the acoustic medium enables us to communicate so that we are able to share our thoughts and intents with other individuals. People rapidly and efficiently encode their intents in speaking, and this fact invites us to think of the acoustic signal as a repository of information about the source of the acoustic signal, the speaking individual. Examined from this perspective, the speech research one finds at meetings of the Acoustical Society of America (ASA) highlights the complexity of such acoustical phenomena. Speech signals are very rapid but are also defined by a complex and intercollated matrix of information. Some of this information tells us about the physics of the person speaking, for example, individuals with shorter mouths and pharynges emit higher frequency resonances than individuals with longer ones. Other parts of this information encode the intent of the individual. This information in speech is governed by a set of shared expectations that the individuals have about how to communicate, that is, this information is governed by their shared language.

Most striking in a perusal of the volumes of abstracts from the ASA meetings over the past 20 years is the large number of contributions examining one particular problem in speech acoustics, second language acquisition (SLA). As SLA research has grown, we have found that the act of speaking is a remarkable feat of motor control. Most speakers are virtuosos in playing the complex scores dictated by our languages, performing rapidly and apparently effortlessly the exacting dictates of many small gestures of the tongue, lips, and larynx. In light of this, multilingualism is even more surprising. It is surprising that individuals can perform these feats of motor control at will and according to the dictates of two different systems of encoding. They have two, very similar sets of skills for rapidly encoding intents that are tailored to different sets of expectations by different language communities.

This article presents some of what has been learned about speech and language through cases in which an individual is not entirely successful at encoding speech in one language because that language was learned later in life and skills were tailored according to the criteria of another language. This variance from what a language group expects goes by a name, *foreign accent*. One other thing to note from the outset is that the foreign accent views the speech signal from the point of view of the listener. A point that is highlighted in this article is just how stringent listeners are; they are often very good at detecting a foreign accent.

Foreign Accent Perception: The Devil Is in the Details?

Traditional research on speech over the last two centuries usually reduces the complexity of the speaking and listening processes by heavily simplifying the speech signal, suppressing much of the information available in the signal. Traditionally, we viewed the signal at the level an alphabetic code similar to western writing systems. This tradition is so pervasive that disciplines related to language and speech science have developed a standard writing system, the International Phonetic Alphabet. This written form of speech strips away a lot of information from the signal. It does so through the process of categorizing bits of the signal into a fairly small number of consonants and vowels.

However, research on perceiving the speech of nonnative speakers amply shows that a language's dictates concerning what a speaker is expected to do is much more finely grained and exacting than a representation by transcription would lead one to believe. Nonnative speakers who are very proficient may, in good listening conditions, produce the basic consonant and vowel categories of a language well enough to convey most of the words in a language. But even so, there is much in the speech signal that is easily detectable by listeners who are native to the community of the language as being different from what is typical of other (native) speakers. Detecting this is termed perceiving a foreign accent, and it turns out that there is a lot of information in the speech signal of the nonnative speaker that can lead listeners to believe that an individual is not a native speaker of the language.

A most striking demonstration of just how picky our expectations for native speech can be was presented at the Miami meeting of the ASA in 2008. Park et al. (2008) reported highlights of a series of perceptual experiments aimed at determining the locus of information that indicates that a particular speaker is someone having grown up in Korea (reported in detail in Park, 2008). The logic of the experiments was to present various portions of speech with a variety of consonants and vowel combinations, thereby controlling the specific linguistic information available to the listeners. On one extreme were clips with fairly complicated two-syllable words, such as "*blanket*," "*breakfast*," and "*razor*," giving many opportunities for speakers to diverge from native productions. Other productions were shorter, progressing down to specific syllables with various combinations of consonants. On the far extreme were productions of just the vowel "ah" ([ɑ], the vowel in the word *"pot"*). Each speaker just produced the vowel by itself. These isolated vowel productions were included to find a minimal case where the native and nonnative speakers converged to be indistinguishable. Then, by comparing different added speech segments, Park (2008) sought to determine which segments were contributing most to the perception of a foreign accent.

Surprisingly, the perceptual responses showed that the native and nonnative productions, even for these isolated vowels, were distinguishable.

These results are explicated in **Figure 1**, which presents measures of *discriminability* for different types of speech. Park treated the perception of a foreign accent as a signal-detection task like the classic signal-monitoring tasks from the early 1950s. See MacMillan and Creelman (2005) for an extensive treatment of how to measure performance in such tasks. Here, the assumption is that a person's nonnative origin is encoded as information embedded in the speech signal and that the listener's task is to detect the presence of that information. The standard measure for such tasks is d', a statistic that evaluates the amount of errors the listeners make, errors in missing the fact that the speaker is nonnative and errors in falsely thinking that a native speaker is nonnative. Average d' values for the experiments are plotted in **Figure 1**. Values of d' of 0 would indicate that the listeners cannot detect the foreign accent. Values of d' for such things as native speakers identifying most consonants produced by native speakers in a quiet environment typically are above 2. Values around 1 indicate, generally, well above chance performance, although not with extremely high accuracy rates.

Not surprisingly, listeners were well above chance at detecting the foreign accent with the two-syllable words. With the one-syllable words, the accuracy rates were similar. What is most striking is that the estimates that speakers were nonnative never converged to the chance level. Even in the cases where the productions included just a single vowel spoken by itself, listeners could still detect some sort of accent in the productions at rates greater than chance.

Various analyses following up on this result showed large differences between the various listeners in their ability to detect the accent in this vowel-only condition. Further analyses failed to reveal any obvious differences in measurements of the acoustics in these cases. Speech science has developed a

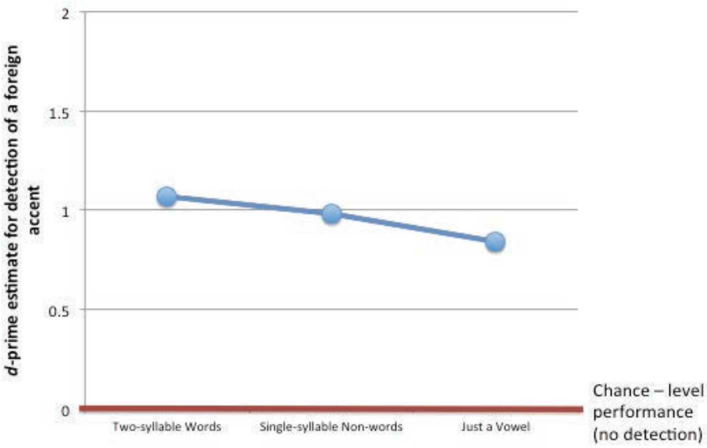

Figure 1. *D' estimates for native listeners detecting whether different kinds of speech samples were produced by a nonnative speaker. The speech samples were produced by English speakers from the United States and nonnative speakers from Korea. Data from Park (2008).*

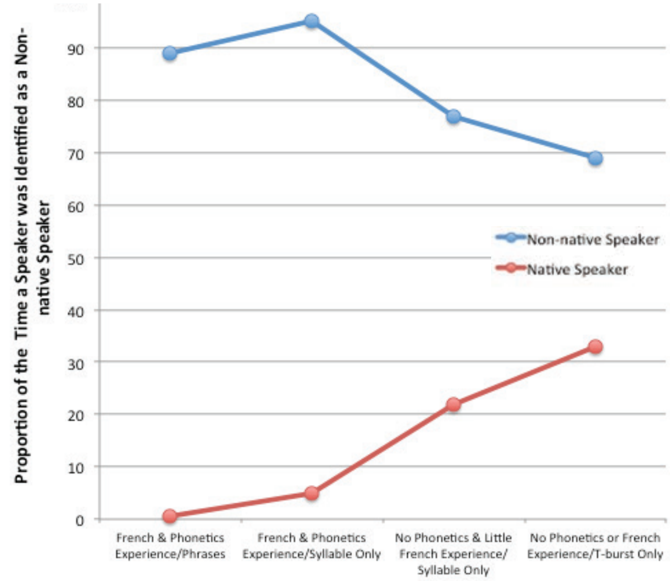

Figure 2. *Proportion of time a stimulus was identified as being produced by a nonnative speaker in four of the experiments reported in Flege (1984).*

number of measurements to quantify many perceptually relevant aspects of speech acoustics, especially those aspects that are known to distinguish words in various languages. However, the level of informational detail indicating a foreign accent is at some finer-grained level than what is used to characterize phonetic behavior with our current instrumentation.

One takeaway from these results might be that vowels are very important carriers of speech information, within which many very fine shadings of relevant information are available. For example, there is a common observation that vowel information is heavily responsible for the differences between English dialects, leading to a well-developed tradition of comparing spectral vowel shapes between speakers of different dialects to map out the geographic variation between the various dialects. There is very extensive research laying out vowel information in a broad range of dialect research, including work published in *Acoustics Today* (Jacewicz and Fox, 2016) and in the enormous volume of material published in the *Atlas of North American English* (Labov et al., 2006). For more targeted research of this type, see Clopper et al. (2005).

However, this kind of fine detail is not just available in vowels. Flege (1984), in a classic article published in *The Journal of the Acoustical Society of America*, reported a series of experiments on the perception of foreign accent in nonnative speakers of English living in Birmingham, Alabama. The native language of the speakers was French. Across the experiments, Flege varied both the content of the speech itself and the experience level of the listeners.

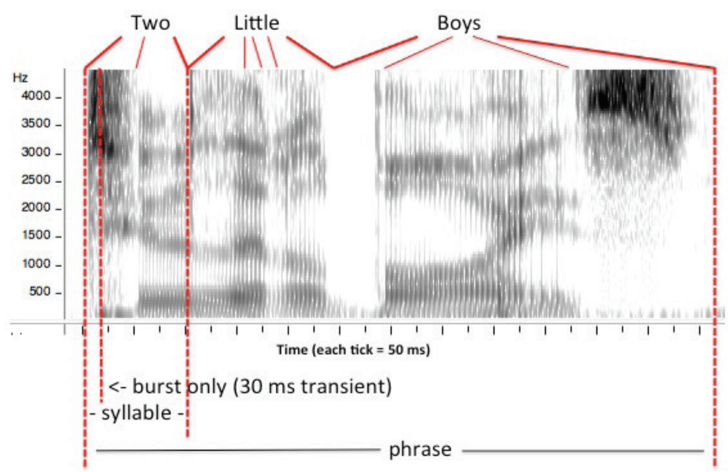

Figure 3. *A spectrographic representation of a native production of the phrase "Two Little Boys."* ***Solid red lines,*** *portion of the spectral image associated with the text at the top;* ***dashed cursors*** *and the three lines of text at the bottom, the three stimulus types that yielded the accent perception rates given in* ***Figure 2****, the largest being the phrase, the next being the syllable, and the final type being just the first 30 ms of the production.*

In the first experiment, Flege (1984) began by having phonetically trained listeners with experience in French listening to entire phrases. Here, he found very high rates of accuracy. Through four experiments, Flege progressively used listeners with less training and presented shorter and shorter bits of speech to these listeners. In the last experiment, he collected phonetically untrained undergraduate students without extensive French experience and presented to them just the burst release from the consonant /t/.

Figure 2 plots how often listeners identified nonnative speakers as nonnative speakers (***blue line***), and native speakers (erroneously) as nonnative speakers (***red line***). As shown in **Figure 2**, accent detection rates were very high with stimuli including phrases, where there are many different opportunities for speakers to diverge from native productions. However, in the final condition (***rightmost circles***), just a portion of one consonant (the stop burst) was still enough to enable listeners to separate native from nonnative speech.

To appreciate just how impoverished the stimuli were, see **Figure 3** that illustrates the sequence of stimuli with spectrographic images of a native production of one of the phrases used, "*Two Little Boys*." The main frame in **Figure 3** contains the complex spectral patterning of an entire phrase. The second and third experiments presented stimuli such as the word *two*, the high-frequency noise of the initial /t/ along with the lower frequency complex of the following vowel between the leftmost and the third red cursor in the panel. In the last experiment, as marked off in **Figure 3, *two leftmost vertical, dashed red lines***, the listeners heard just the 30-ms bit at the very beginning. This bit contains a portion of what is called a burst release, which is the noise associated with the small puff of air that occurs when speakers open their mouths in the production of some consonants such as /t/. Acoustically, this is a short, noisy transient not more than 30 ms in duration. The foreign accent detection rates, even in this extraordinarily impoverished condition, were well above chance (see the difference between the ***rightmost red*** and ***blue circles*** in **Figure 2**).

The overall point of these research findings is obvious. If you want to produce speech in a new language that is indistinguishable from that of native speakers, you have to set a very high standard. There are very tight criteria involved, not just with the production of each vowel but also with the production of many aspects of the consonants. Listeners are very accustomed to the speech of their community, and even very small divergences from this speech can be detected by the listeners. Also, while not all speech sounds necessarily exhibit appreciable divergences across languages, a good many do, and research on foreign accent has detected many such cases. Where the limits of detectable foreign accent lie, in terms of individual speech sounds, still has not been fully determined.

Problems of a Higher Order

Even as discouraging as this might be for a language learner, learning to produce second language speech is actually even more difficult. Differences between languages not only are found in the consonant and vowel pieces that make up speech but also in the higher order structure of the speech. Speech is not just a string of individual acoustical letters but involves the whole orchestration of these bits into the larger complex signal. Foreign-accented speech, then, also involves the dynamics of the acoustical patterns that arise in the sequencing of consonants and vowels.

To illustrate this point, I turn away from studies specifically of foreign accent detection to work that has been done on intelligibility. Intelligibility refers to measures of how likely a person is able to identify what a person is saying in a recording.

A particularly striking example of this problem of higher-order organization in second languages was demonstrated by a sophisticated study of the effects of dynamic modulations in nonnative speech on intelligibility by Tajima et al. (1997). Tajima et al. were puzzling through on how to deal with the fact that speech patterns unfold in time in many ways. The speech signal does not have a fixed timing pattern, but timing is modulated by many well-known factors. In the process, they worked with a technique called dynamic time warping, a technique whereby one could take two related acoustic signals with different timing patterns and determine a mapping between them. See, e.g., Rabiner et al. (1978) for an early speech foray into dynamic time warping. From this mapping, using various computational techniques, one can hybridize the two signals into one with the spectral patterns from one signal and the timing patterns from the other signal.

This mapping pattern from one production to another is illustrated in **Figure 4**, which is an example from Tajima et al. (1997). The spectral images here have the durational patterns of two productions of the phrase "*shelled egg,*" a foreign-accented production at the top and an unaccented one at the bottom. Tajima et al. took the timing pattern from nonnative Chinese productions of this and other English sentences and hybridized them with the spectral patterns of native productions of the same sentence. The outcome of this process, then, had the spectral properties of the native speech but the timing patterns of the nonnative speech. He then presented these hybridized recordings to native English listeners with various levels of masking noise and had them write down what they thought the person was saying. The logic of the study was to determine the effect of nonnative timing patterns on intelligibility. As a parallel test, he also took the timing patterns of native productions and hybridized them with the spectral properties of nonnative productions, thereby "correcting" the nonnative timing patterns.

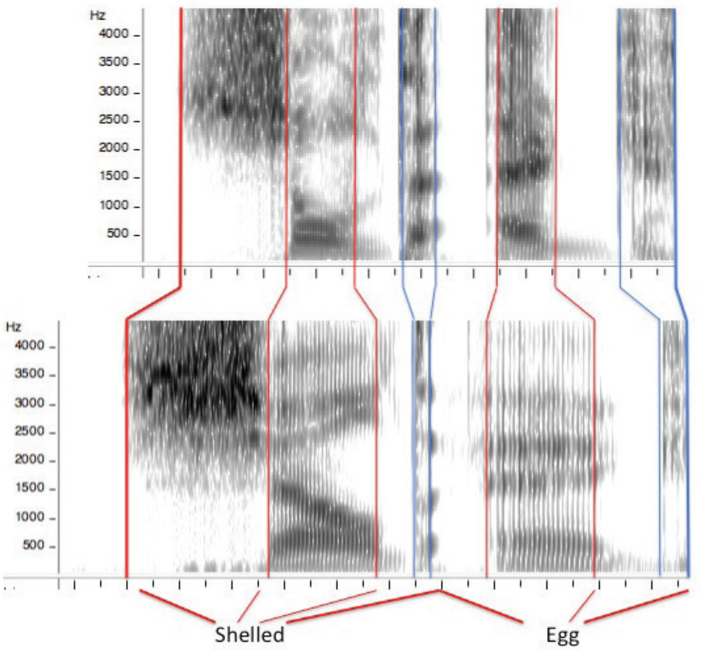

Figure 4. Spectrographic images of two productions of the phrase "shelled egg," *patterned after productions examined in Tajima et al. (1997).* ***Top:*** *duration patterns from a nonnative speaker of Mandarin Chinese;* ***bottom:*** *duration patterns from a native speaker of American English.* ***Angled lines,*** *portions of the spectral images associated with the text at the bottom;* ***red lines,*** *mapping of timing patterns between the two productions, by connecting parallel segments;* ***blue lines,*** *"extra" segments in the nonnative productions, where there is a strong, vowel-like release to the last consonant in* "shelled" *and* "egg."

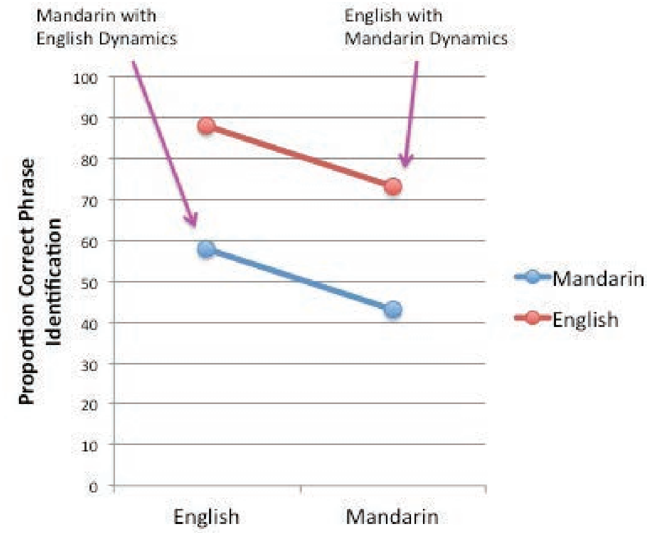

Figure 5. *Estimates of intelligibility for native speakers of Mandarin and of English, with timing patterns imprinted from matched English and Mandarin productions. Data from Tajima et al. (1997).*

The intelligibility evaluation scores, shown in **Figure 5**, tell the story of the effect of the hybridization. Native productions with native timing patterns (***top left red circle***) yielded intelligibility estimates around 90%, and nonnative productions with nonnative timing patterns were much less intelligible, about 40% (***bottom right blue circle***). Once these timing patterns were artificially corrected, the speech of the hybrid nonnative speech with the "corrections" from the native speech was substantially more intelligible, rising almost 20%. These intelligibility offsets went in both directions so that native speech, with almost 90% phrase accuracy in noise, lost nearly 10% when hybridized with the timing patterns of the nonnative speaker.

The point of these results is fairly clear: there is more to speaking the second language as a native than just getting the individual consonant and vowel qualities right. The whole ensemble of consonants and vowels needs to be executed with the right timing patterns to be treated as native speech.

A closer look at Tajima et al.'s (1997) results suggests how disastrously unintelligible accented speech can be when the speech occurs with background noise. (The data from **Figure 5** were obtained by embedding the speech with a −5-dB noise masking.) It also points out a particularly troublesome locus for the listeners' errors, pointing to yet another aspect of different languages that presents a challenge to the nonnative speaker. Many of the most troublesome errors occurred in cases in which the hybridization process had to either add or delete an acoustical segment completely. For example, as illustrated in **Figure 4**, one of the phrases was "*shelled egg*," in which the native speakers neatly run the final 'd' in "*shelled*" into the vowel at the beginning of "*egg*." Chinese native speakers, by contrast, produced a very robust consonant release for the 'd' in "*shelled*." This nonnative production was heard by listeners as various things, including the phrase "*shall I ask?*" Although the intended form only has two syllables, "*shelled*" and "*egg*," the perceived form ended up with three syllables: one, apparently the "*I*" corresponding to the strong release of the 'd' consonant. There were quite a number of these sorts of errors, including, for example, the intended "*limit contour*" being heard as "*leave it on the counter*," with "*on*" corresponding to the release of the 't' in "*limit*" and the intended "*change color*" being heard as "*twenty caller*," with the 'ty' on "*twenty*" corresponding to the 'ge' on "*change*."

The reason for this heavy releasing effect seems to be related to the possible sequences of sounds in Mandarin Chinese and English. English allows words to end with consonants, such as the last one in "*limit*" and "*change*," but Mandarin does not. Different languages allow different sequences. Some languages, such as Japanese, are extraordinarily restrictive, traditionally allowing only words with a single consonant at the beginning (with some consonants requiring an additional vowel-like articulation such as 'ky' /kj/) and no consonants except, possibly, a nasal segment such as 'm' or 'n' at the end. This differs from English, which also allows consonants to combine into some fairly complex clusters, both at the beginning, such as in the word "*spline*" beginning with /spl/, and at the end, as in the word "*sixths*" ending with /ksθs/. Actually though, even with these allowances, English is not all that extreme in the proliferation of consonant clusters. Languages from around the globe, such as Georgian (a language of the Caucasus Mountains), Berber (a language of North Africa), and Nuxálk (a language of the North American Pacific Northwest) regularly exhibit much larger consonant sequences, allowing extraordinarily long sequences of 6, e.g., Georgian /prt͡skvna/ (meaning "*peeling*"), or even more consonants, such as the commonly cited 12-segment Nuxálk word /xɬp'χ^{w}ɬt^{h}ɬp^{h}ɬːsk^{wh}t͡s'/ (meaning "*He had had in his possession a bunchberry plant*"), which is made up entirely of consonants.

As can be readily imagined, these sequencing differences create many problems for speakers of a language like Japanese learning to speak a language such as English (or an English speaker learning Nuxálk!). Not only do the various speech actions have to become coordinated with one another but also the combinations of actions themselves begin to take on aspects of the productions that characterize the combination as a whole. These sequencing difficulties are well researched. Excellent examples of this come from a wide range of studies on the particular problems created for learners whose native language is very restrictive of consonants following a vowel, such as Japanese, Korean, and Italian (a fact reflected in the fact that the word "*Italian*" could not appear in Italian without a final *o*).

In these cases, it is common for speakers of various languages to initially simply fail to produce a consonant with enough of an acoustic signature for listeners to hear it. Speakers with more experience tend to go in another direction, noticed by Tajima et al. (1997), of overarticulating the consonant, perhaps to ensure its being heard. The outcome of doing this is the perception by native listeners of an additional vowel following the consonant, as was found pervasively in Tajima et al.'s study. As suggested in Tajima et al.'s study, these sorts of overproductions can actually be particularly harmful for intelligibility. Although one might get most of the consonant and vowel identities correct, the addition of perceived vowels can push a listener to add various endings on the word to account for the additional vowel or to add short function words such as "*the*" or "*in*." This, in turn, can often force the interpretation of the misperceived word as in the wrong grammatical category, for example, hearing a noun as a verb, which will then force a completely different reading of the overall sentence to make sense of the location of the verb where a noun was intended.

Applying the Criteria: What People Do with Accented Speech

All of this sensitivity to variance in the language by listeners is generally bad news if you are trying to communicate as a native speaker. There is, however, some good news in the SLA research on accent perception for learners. Various studies also show that although listeners often pick up on the deviation from the native patterns, listeners can, and often do, accommodate to someone who they think is a nonnative speaker.

An interesting demonstration of this point was found in a study by Munro and Derwing (2001). Munro and Derwing were interested in understanding a very pervasive overall timing difference between nonnative and native speech. Native speech is typically spoken faster than nonnative speech. To study this, they developed a technique of digitally compressing or stretching various portions of speech recordings to produce artificially faster or slower speech. They then used the artificially manipulated the speech to determine how people react to the fact that nonnative speakers are typically slower than native speakers.

They presented the speech to listeners and asked them to evaluate the modified speech for nativeness. What they found is that not only are nonnative speakers slower in speed but also that native listeners then take the slower speech and evaluate it as less native than faster speech. So this is similar to the studies just reviewed already; all aspects of nonnative speech, apparently, can cause listeners to identify it as nonnative, even something as global as the overall speed of production.

However, there is a twist in Munro and Derwing's (2001) results. They also found that speech that is unnaturally fast was *also* evaluated as less native, so listeners had a preferred speed of production somewhere in the middle. From these data, then, they could estimate an optimal rate of speed for making the speech most native-like, and this estimate had an interesting property. While being faster than the natural speed of the nonnative speakers, it is still not as fast as that found in native speakers. That is, it appears that while nonnative speakers are generally slower than native speakers, to some extent, the listeners expect it and like it that way.

This sort of higher level accommodation to nonnative speakers is quite pervasive, and, although it is not all benign, it often can be quite helpful. A good example of this accommodation was found in another, very different sort of study of language learners. This study examined how learners of a second language can perform in expressing sarcasm. Alexander (2011) created a number of different role-play scenarios in which speakers would perform an individual apologizing for various reasons. She also devised a number of other scenarios in which the speakers would role-play an individual *not* apologizing, producing speech with markers of sarcasm.

For example, in one scenario the speaker was put into a situation where there was an accident in which they spilled hot coffee on a conversation partner who was a stranger to them. In such a case, the clear expectation was that a person saying "*I'm really sorry…*" is actually apologizing for the accidental damage to the person of the interlocutor. In a contrasting scenario, the damage to the interlocutor is just a nontangible offense taken by a roommate for something that the speaker does not think is reasonably something to give offense. In this case, saying "*I'm really sorry …*" is not an actual apology, and the speakers were to convey this.

Alexander (2011) then had both native American speakers and nonnative speakers from Thailand, either living in Thailand or America, perform the role-play scenarios. In addition to analyzing the acoustics of the recordings of these sincere and sarcastic apologies, she took clips from the recordings of individuals saying "*I'm sorry*" and the like, presented them to native listeners, and asked them to decide if the speaker was being sincere or sarcastic.

Alexander (2011) used a similar statistical treatment as Park (2008), treating it as a signal-detection task. Here, the signal to be detected is an apology. The sarcastic apologies, then, are considered nonsignals which the listeners are to correctly reject as a nonapology.

Figure 6 plots the summary data from the study. The listeners exhibited d' values for the native productions that were above chance. Native speakers could detect the sincerity in these apologetic clips while rejecting the insincere apologies. Also, a second statistic (c; **Figure 6,** ***red circles***) measures the degree

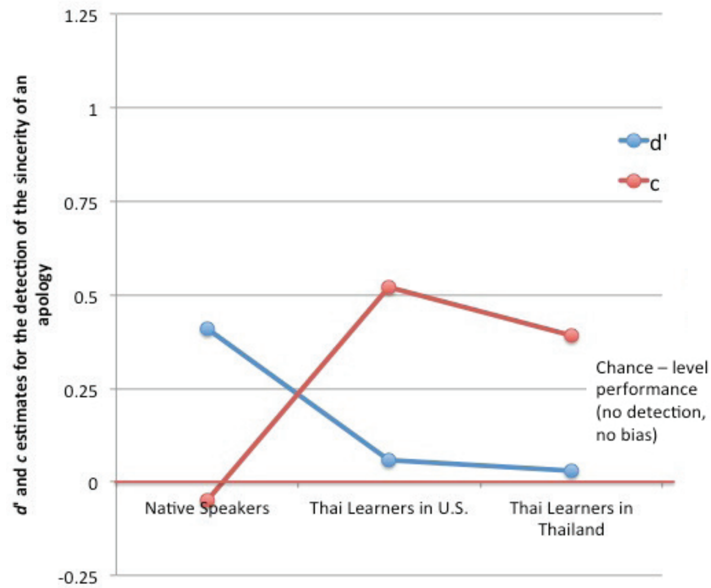

Figure 6*. Blue line,** *d' values for native listeners of English detecting whether an apology is sincere for different speaker groups;* ***red line, *measure of the degree to which responses are biased to one response (c).*

to which the listeners were biased toward thinking all of the apologies were sincere or insincere. The values of c for the native speakers are very near zero, indicating very little response bias one way or another. This suggests that they were very accurately tuned to pick up on the sarcasm of the responses.

The nonnative speakers, however, did not fare so well; d' values were very near zero, indicating that listeners were very bad at being able to detect whether the nonnative speakers were trying to be sincere or not. However, here there is a bright, silver lining to the cloud. Analyses of c indicate a very strong bias in the listeners toward saying that the nonnative listeners were sincere. That is to say, the listeners were strongly disposed to treat any apology, sincere or not, as sincere if the nonnative speaker produced it. Note that this was not the case with the native speakers, where the decisions were much more attuned to the marks of insincerity in the recordings.

So, although there may be many demonstrations of people's negative adjudication of foreign-accented speech, where there is detection that someone is a nonnative speaker, there is also, then, the possibility that a listener can explain divergences from what is expected as being due the speaker being nonnative. It is part of how we make allowances for nonnative speakers. In the case of Alexander's (2011) work, this appears to explain why native listeners were much more willing to take the nonnative apologies at face value. For nonnative speakers, this is probably a good thing; the production differences due to our nonnative experience can and often are accounted for by the listeners.

Conclusion

As an acoustical phenomenon, the speech signal is a remarkably rich and complex repository of information about an individual's intent in speaking. But it is also filled with information about the speaker. We are remarkable in our ability to perceptually evaluate this rich and complex repository, not only in decoding the complex messages that the speaker intended but also in evaluating the similarity of the speech patterning to our native community. In the case of nonnative speakers, we often find just how stringent the criteria for native-like speech is and thus how detailed the acoustic information relevant to us as humans is.

References

Alexander, S. T. (2011). *Sincerity, Intonation, and Apologies: A Case Study of Thai EFL and ESL Learners.* PhD Thesis, Departments of Second Language Studies and Linguistics, Indiana University, Bloomington, IN.

Clopper, C. G., Pisoni. D. B., and de Jong, K. J. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *The Journal of the Acoustical Society of America* 118, 1661-1676.

Flege, J. E. (1984). The detection of French accent by American listeners. *The Journal of the Acoustical Society of America* 76, 692-707.

Jacewicz, E., and Fox, R. A. (2016). Acoustics of regionally accented speech. *Acoustics Today* 12(2), 31-39.

Labov, W., Ash, S., and Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology and Sound Change.* Mouton de Gruyter, Berlin.

MacMillan, N. A., and Creelman, D. C. (2005). *Detection Theory: A User's Guide,* 2nd ed. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Munro, M. J., and Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition* 23, 451-468.

Park, H. (2008). *Phonological Information and Linguistic Experience in Foreign Accent Detection*. PhD Thesis, Department of Linguistics, Indiana University, Bloomington, IN.

Park, H., de Jong, K. J., and Darcy, I. (2008). The amount of information needed for listeners to detect a foreign accent. *The Journal of the Acoustical Society of America* 124, 2593.

Rabiner, L. R., Rosenberg, A. E., and Levinson, S. E. (1978). Considerations in dynamic time warping algorithms for discrete word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26, 575-582.

Tajima, K., Port, R., and Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics* 15, 1-24.

Biosketch

Kenneth de Jong is professor of linguistics and cognitive science and adjunct professor of second language studies at Indiana University, Bloomington, IN, conducting research into the diversity of processes pertaining to speech production and perception from motor coordination to the acoustic structure of linguistic categories to the acquisition of perceptual abilities in a second language. He was editor in chief of the *Journal of Phonetics*, is currently president of the Association for Laboratory Phonology, and was cochair of the 168th meeting of the Acoustical Society of America in Indianapolis, IN.