Marnie Reed is an Associate Professor in the graduate TESOL Program and the Program in Applied Linguistics at Boston University. The focus of her research has been on second language acquisition, with a focus on the relationship between speech production and perception. She has co-authored the books *Sound Concepts: An Integrated Pronunciation Course* (2005, McGraw-Hill) and its accompanying *Teacher’s Manual* (2006), and *Goal-Driven Lesson Planning for Teaching English to Speakers of Other Languages* (2010, University of Michigan Press), which incorporates her work on pronunciation instruction.

Christina Michaud is a Lecturer in the College of Arts and Sciences Writing Program at Boston University and also a doctoral candidate in Language, Literacy, and Cultural Studies in the School of Education at BU. She teaches ESL composition as well as general first-year composition and research writing. She has co-authored two books and numerous conference presentations with Marnie Reed.

An Integrated Approach to Pronunciation:

Listening Comprehension and Intelligibility in Theory and Practice

Abstract:

This paper introduces a theoretical basis for understanding the relationship between speaking and listening as an auditory feedback loop, in which speakers use their own mental model of a sound as input. We argue that speaking and listening are related: production facilitates perception, promoting more intelligible spontaneous speech and enhanced listening
comprehension. Since learners use their own speech production as a filter, we suggest a model based on convergent production. We compare our model to Flege's Speech-Learning Model and also apply our model to the practical problem of pronunciation instruction.

We identify four major components of pronunciation instruction: connected speech features, suprasegmental features, inflectional morphology, and segmentals. Within our integrated model of pronunciation, the route to successful listening comprehension is through auditory feedback wherein the learner's own increasingly target-like speech production facilitates and reinforces perception. We introduce specifics for promoting students’ convergent output: spoken models, visual aides, oral or written descriptions, etc. We advocate a teacher-student partnership in which teachers offer a principled approach to corrective feedback in the form of production prompts. In our model, pronunciation instruction accompanies and reinforces core language instruction. We view integrated pronunciation instruction as a highly focused, metacognitive approach to the entire language classroom.

Paper:

Data and anecdotal observations alike indicate that adult second language learners often have intractable pronunciation problems that persist in the face of explicit instruction or correction. Mere input of the target sounds is insufficient (Flege & Hillenbrand, 1984; Flege, 1993; Strange, 1995), obviously, or our learners would never need instruction. The long-established failure of adult learners to incorporate syntactic structures to which they are exposed (Pavesi, 1986; Long et al, 1998) has a parallel in pronunciation. Students frequently seem not to “hear” the target pronunciation even when modeled by their teachers, instead immediately continuing with their incorrect pronunciation.
One way of making sense of this problem is to conceive of the relationship between a learner’s speaking (production) and listening (perception) as an auditory feedback loop, in which speakers use their own output—their own mental model of a sound—as input for their production. The role of auditory feedback, attested in research on speakers with normal hearing and profound hearing loss (Perkell et al, 1997, 2000) also helps account for the persistence of both spoken error and a lack of aural discrimination between the target and the error.

The question then arises of how to break this feedback loop; naturally enough, it can be broken when a speaker actually forms the target sound in his or her mouth and then combines this new and different motor-memory with a new acoustic image of the sound. In practical terms, we tell learners that when they finally “get it right”—produce a target sound—it will “feel wrong” to them, but they should nonetheless perform a mental freeze-frame at that moment.

For our purposes, it doesn’t matter at all how to get speakers to produce the sound: teachers may use any means at their disposal, including aural models; mouth diagrams and mirrors; metalinguistic feedback; and pronunciation aids involving rubber bands, Chinese yo-yos, kazoos (Gilbert, 2005; Grant, 2004, Hewings, 2004). We tend to think, based on the research (Couper, 2006; Gilbert, 2010), that aural models will be less helpful, by and large, than other means, but teachers should on no account rule out any particular way of getting learners to produce the target sound.

Often speakers *can*, in fact, produce the target sound (Fraser, 1999, 2000), but in limited phonological environments. Word-initial /w/ in “want or “what” is not problematic for Japanese speakers; however, since upsilon does not occur in Japanese, word-initial /w/ in “would” or “woman” is difficult. By helping the speaker to isolate the target sound in a licensed consonant-
vowel combination, we can start to help the speaker to form a new mental model in expanded c-v contexts.

Once teachers can get a learner producing the target sound, in any context, and on the road to creating a new mental model of it, both the learner’s speaking and listening stand to improve. We believe that in a model of convergent production, production precedes and facilitates perception, promoting more intelligible spontaneous speech and enhanced listening comprehension as the speaker’s production begins to converge on the target: as we tell students, speaking helps listening. We posit that our production-centered approach to intelligibility is actually consistent with Flege's Speech-Learning Model (1995). Flege’s SLM focuses on the mental and metacognitive work of perceiving a new target sound; our model does not contradict the SLM, but instead offers a specific and practical pedagogy to allow learners to create acoustic images of new sounds.

With what we see as this key relationship between speaking and listening in mind, therefore, we argue for an expanded model of pronunciation which includes both speaking and listening and addresses both suprasegmentals and segmentals. Figure 1, below, represents our Integrated Model of Pronunciation:
We identify four major components of pronunciation. Working from left to right in the figure above, we identify the first strand as listening for content, because learners who struggle with this strand of pronunciation tend to miss or misunderstand the actual content of what is being said. This strand contains the connected speech features—linked, reduced, deleted, altered, and contracted sounds. Though learners often think that the speed of English accounts for their listening problems, in fact these connected speech features are the contributing sources. One of these features, reductions, has been frequently cited as a major contributing factor to misinterpretation of the homophones 'to' and 'two' in civil aviation (Waldock, 1994), resulting in a 1989 crash when an air traffic controller cleared an aircraft to descend to "two four zero zero" but the pilot interpreted the number 'two' as the preposition 'to' and descended accordingly (Cushing, 1995).

Next, we identify the strand of intent—the suprasegmental features of English (syllables, stress, intonation, rhythm, and timing) that convey speakers’ meaning. Learners who struggle with this aspect of pronunciation may understand every word a speaker says but may still miss the meaning; conversely, they may be unable to get their own content or intent across. This is why we show a dotted arrow coming down to this strand from speaking, as well; we argue that suprasegmentals are essential both for learners’ speaking and listening (Hahn, 2004). In a study currently being undertaken, advanced-level students (D level in an A-E level intensive English program) were questioned prior to suprasegmental instruction on their opinions regarding the relative importance of segmentals versus suprasegmentals, and the role of one component of suprasegmentals: intonation. Using a digital student response system (clickers), students read the following questions from a power point slide and voted their choice for the response:
1. Which is more important for comprehension?
   a. Producing correct consonant and vowel sounds
   b. Producing correct stress patterns in words and phrases

Consistent with the findings of Derwing and Rossiter (2002), more respondents voted for the segmentals: 58% voted for producing correct consonant and vowel sounds while 42% believed that correct word and sentence-level stress patterns were more important.

2. Which is more important in a sentence:
   a. the words in the sentence
   b. the intonation in the sentence

For this question, 73% of the respondents voted for choice a: words, while 27% of the respondents said that intonation was more important. Following completion of the first lesson on intonation in their pronunciation text (Grant, 2009), students were asked their interpretation of a sentence with non-standard sentence-level stress. Using the digital student response system (clickers), students voted their response after hearing an audio recording of a sentence.

   Audio Prompt: *The teacher* didn’t grade your papers.
   Question: Were the papers graded?
   a. Yes
   b. No

The votes split evenly, seemingly reflecting some effects of instruction and generating much discussion. When the correct response, (a), was displayed, a number of students asked to hear the sentence again. Several could be seen and heard repeating the sentence under their breath, particularly the verb phrase "didn't grade." These respondents demonstrated grasp of the lexical, morphological, and syntactic content of the utterance, but missed the intended meaning.

The third strand from the left, under our umbrella term of pronunciation, is here identified as verb and noun endings—in other words, the sounds of English morphosyntax, specifically bound inflectional morphology, which cause students serious and intractable problems when they
omit or mispronounce these endings. Teachers, based on students’ ability to state the grammar rules and produce these endings in controlled drills and tests, appear to succumb to what Gass and Varonis (1985) refer to as accommodation, not noticing the errors in students’ spontaneous speech. The fact that students fail to pronounce these endings on regular nouns and verbs when reading aloud convinces us that the problem is not a local grammar error, but is based in their internalization of the sound system. Studies have shown that errors with these endings — although to students’ ears, small, and sometimes not even the sole conveyance of grammatical meaning in a construction—nevertheless stigmatize learners and are responsible for communication breakdowns (Jiang, 2007; Lardiere, 1998; Long, 2003; Major, 1995, 1998). The misinterpretation of “He looked it up” as “He looked up” by a speaker who pronounces ‘look’ + past tense as two syllables attests to the role of the acoustic image and the interrelation of production and perception. For these reasons, we believe that this strand, too, is key for both learners’ speaking and listening skills, and have tried to represent this on the diagram above.

Finally, our fourth strand of pronunciation is that of individual consonant and vowel sounds—segmentals—which students typically report to be their biggest pronunciation errors and main source of communication breakdowns (Derwing & Rossiter, 2002). Our experience in a pronunciation elective class with a Japanese neurosurgeon who did not differentiate between liquids /l/ and /r/ or fricative /j/ and affricate /z/, rendering 'brain regions' indistinguishable from 'brain lesions,' attests to the need to address segmentals. Nevertheless, we weight this strand as relatively less important, in the grand scheme of intelligibility, than the two strands toward the middle of the figure. Our model advocates a balanced focus on segmental and suprasegmental features, and an integrated approach to instruction.
We believe that this integrated model of pronunciation is particularly helpful to learners for making sense of their errors: teachers are able to break the large topic of “pronunciation” down into meaningful, yet discrete chunks, and then build it back up as learners master various patterns and sounds. Teacher tools that can help learners have success with these different parts of pronunciation abound: we suggest a teacher-learner partnership, where teachers guide learners through both the production of these different pronunciation topics and the metacognition necessary for “noticing” these patterns (Schmidt, 1990) and building up new acoustic images of them (Neufeld, 1977).

Three sample entries in a pronunciation logbook—along with directions written for students—follow, in Figure 2, showing one way that metalinguistic feedback can help move a learner toward more target-like production—in this case, of the problem of initial consonant clusters, a syllable structure problem we classify in the second strand (suprasegmentals, or intent/rhythm) above:

<table>
<thead>
<tr>
<th>Word:</th>
<th>How should I say it?</th>
<th>How did I say it?</th>
<th>What was my mistake?</th>
<th>Other examples:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student 1:</strong> speech</td>
<td>speech (1 syllable)</td>
<td>su-peech (2 syllables)</td>
<td>separating the first two consonants</td>
<td>su-trong/strong</td>
</tr>
<tr>
<td><strong>Student 2:</strong> speech</td>
<td>speech (1 syllable)</td>
<td>es-peech (2 syllables)</td>
<td>adding a vowel sound at the front</td>
<td>es-port/sport</td>
</tr>
<tr>
<td><strong>Student 3:</strong> speech</td>
<td>speech (1 syllable)</td>
<td>speech-ee (2 syllables)</td>
<td>adding a vowel sound at the end</td>
<td>each-ee/each</td>
</tr>
</tbody>
</table>
In Figure 3, below, we offer a checklist for learners who “know” the rule for the pronunciation of the regular past-tense verb ending, but nevertheless omit it entirely or incorrectly pronounce it as two syllables even in verbs like “looked” and “used.” These errors, of course, relate to the third strand (verb and noun endings, or morphosyntax) in our integrated model of pronunciation:

<table>
<thead>
<tr>
<th>Checklist: How do you say the “-ed” ending on regular past tense verbs?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Look:</strong> Find the simple (root) form of a verb, without any endings.</td>
</tr>
<tr>
<td><strong>Ask:</strong> What is the final sound (not letter)?</td>
</tr>
<tr>
<td><strong>Is it:</strong> /t/ or /d/?</td>
</tr>
<tr>
<td>If yes... ✓</td>
</tr>
<tr>
<td>→ Add an extra syllable.</td>
</tr>
<tr>
<td>→ Say “-ed” as [Id].</td>
</tr>
<tr>
<td>If no...</td>
</tr>
<tr>
<td><strong>Ask:</strong> Is the final sound unvoiced?</td>
</tr>
<tr>
<td>If yes... ✓</td>
</tr>
<tr>
<td>There is no extra syllable.</td>
</tr>
<tr>
<td>→ Say “-ed” as [t].</td>
</tr>
<tr>
<td>If no...</td>
</tr>
<tr>
<td>There is no extra syllable.</td>
</tr>
<tr>
<td>→ Say “-ed” as [d].</td>
</tr>
</tbody>
</table>

We argue that teacher intervention is key to helping learners move from isolated production of a new pronunciation sound or pattern to full production of it in spontaneous speech, and in turn to full perception of it in other speakers’ connected speech. Such intervention can in some way be seen as the real work of the pronunciation classroom—it is what teachers are there to do, after all, and it is why we don’t advise students just to improve their pronunciation by merely reading a pronunciation book (even a good one). We describe the intermediate stage when learners are able, sometimes, to produce a target sound or pattern on demand, but have not
yet integrated it fully into their spontaneous speech as the stage of prompted production: through instruction in the sound or pattern, and careful prompting (again, the actual means of getting the learner to produce this sound may vary widely), learners will achieve their first success at producing the sound in a target-like fashion. Learners then move through a stage when teachers’ corrective feedback is necessary to help them solidify the pattern in their own mouths and minds, before they fully form a new mental model of the sound or pattern. To better help learners, we suggest that the language of instruction match the language of correction; in other words, whatever metalinguistic feedback teachers wish to offer (“Say it as one syllable,” “Don’t use a voiced sound there,” “Make the stressed syllable louder, longer, clearer, and higher,” etc.) in the classroom, to individual learners, should be the same terms as those in which they teach the concept or pattern to begin with. That language should also be the same that teachers elicit from students in the form of student “tell-backs” (a term we borrow from the literature on reading instruction [Vanderwood, 2007])—student language that articulates, superficially for the teacher’s benefit, but more significantly for the learner’s own benefit, what the learner now knows about this concept (“I said it as two syllables, but I need to only say one; I need to omit the final syllable after the sibilant.”) and what will help the learner go about forming a new mental model. Over time, as students demonstrate internalization and metacognitive awareness, external corrective feedback can take the form of increasingly precise, minimal, and unobtrusive 'production prompts' which allow the learner to self-correct and return to finish their thought. Figure 4, below, shows what we consider the path to intelligibility, and the connections among the different terms we have been using in this model:
In our model, pronunciation instruction accompanies and reinforces core language instruction. We view integrated pronunciation instruction as a highly focused, metacognitive approach to the entire language classroom, and a key step along the way to learners’ increasing intelligibility.

References


