Remediation of /r/ *for* Speech-Language Pathologists

Peter Flipsen Jr., PhD, S-LP(C), CCC-SLP

PLURAL PUBLISHING INC.

CHAPTER 2

Typical Production of American English /r/

Given the challenges so many clinicians appear to be having with this sound, a refresher and/or update on how /r/ is produced seems appropriate. If we really want to sort out the problem, it makes sense to have a precise understanding of the goal. Although some readers will be tempted to ignore this chapter, I suspect that at least some of what is about to be presented here will be new to most.

ARTICULATORY FEATURES OF /r/

Consonant sounds are typically described relative to voicing, manner of articulation, and place of articulation. There appears to be little controversy regarding voicing and manner for American English /r/. It is generally accepted to be a voiced sound, although the possible loss of voicing in some contexts has been suggested. Its manner of articulation is listed variously as a *rhotic*, an *approximant*, or a *liquid*, and all three terms are quite appropriate. According to Ladefoged and Maddieson (1996), American English /r/ is one of several sounds belonging to the category *rhotic*. This category is unusual in phonetics in that it simply refers to those sounds that happen to be represented in spelling, when using the Latin (Roman) alphabet, by the letter "r." The terms *approximant* and *liquid* are more directly descriptive of the manner of articulation for /r/. They are somewhat synonymous terms in that both refer to a vowel-like sound in which there is only slightly more constriction in the vocal tract than for most vowels.

Three Places of Constriction (Articulation)

Adequately describing /r/ becomes complicated when defining its place of articulation. As with most speech sounds, it is the location of the narrowing or constriction for /r/ that yields its particular sound quality. As mentioned in Chapter 1, /r/ includes three different places of constriction rather than one or two seen with other sounds. It is likely that in your phonetics training, you only discussed two constrictions with the main focus for /r/ being around the constriction in the oral cavity; typically it is said to occur at the palate. However, even that description may be inadequate. If you consider the two most commonly mentioned tongue shapes for /r/ (retroflex and bunched; to be discussed later), there appear to be two somewhat different places where a narrowing of the oral cavity would be occurring: (a) near the front of the palate (just behind the alveolar ridge) for a retroflex tongue shape or (b) farther back on the palate (just in front of the velum) for a bunched tongue shape.

The second constriction that only gets passing mention in most phonetics classes is at the lips. American English /r/ is typically described as involving lip rounding (somewhat less than that seen for most English back vowels), although the amount of lip rounding for /r/ can vary greatly from one speaker to the next and from one word context to the next.

The third constriction for /r/ is often not mentioned in most undergraduate phonetics classes. This is a constriction in the pharyngeal cavity involving some degree of narrowing between the tongue root and the posterior pharyngeal wall. Although there is still not much understood about this constriction, given that at least two different tongue positions (shapes) yield the same acoustic output, at the very least speakers likely make simultaneous adjustments to both the oral and pharyngeal constrictions. This is suggested by Boyce (2015) who noted that "the pharyngeal constriction in American English speakers tends to be narrower for tongue configurations with a raised tongue tip [retroflex] and wider for tongue configurations with a 'bunched' configuration" (p. 262).

The need for and the interaction among the constrictions was also noted by Delattre and Freeman (1968):

> Experiments with an electronic analog of the mouth, which permits the shifting of constrictions and the observation of the auditory effect of each modification, have shown that the pharyngeal cavity alone does not produce an American /r/ . . . it is the palate-velar constriction which produces the American /r/ . . . as a constriction is slowly moved from the alveols [sic] toward the back, the auditory impression of the American /r/ increases, reaches a maximum near the frontier of the palate and velum and rapidly disappears beyond that point. When the palato-velar constriction is held, if the pharyngeal constriction is

> narrowed, the auditory impression of /r/ is enhanced . . . if the pharyngeal constriction is widened, the /r/ is subjectively mellowed but does not disappear. (p. 42)

Ultrasound is now providing additional insight on this issue. In an ultrasound treatment study, Preston, Leece, and Maas (2017) reported that "(c)orrect productions were generally associated with elevation of the anterior tongue and depression of the tongue dorsum indicative of tongue root retraction" (p. 86). Dugan and colleagues (2019) also used ultrasound with four children to conduct a detailed analysis of movements of different parts of the tongue; they reported that the two children with normal /r/ productions used more tongue movement overall and moved individual parts of the tongue to greater degrees than the two children who produced /r/ errors. Even if ultrasound intervention remains beyond the reach of most of our patients and families, such analyses may be quite valuable. They are beginning to provide us with cuing strategies to provide to other patients (i.e., those without the benefit of ultrasound feedback) to teach them to modify those constrictions and generate acceptable /r/ productions.[1] These are discussed later.

It is in fact likely that English speakers are coordinating all three constrictions to produce a correct /r/. You may recall from your acoustics class that American English /r/ is characterized by a lowering of the value of the third formant (F3). *Perturbation theory* suggests that lowering of F3 results from a combination of all three constrictions (lips, palate, pharynx). Lindau (1985) used a combination of acoustic and x-ray data from six speakers of American English to reach a similar conclusion, noting that "it seems that speakers of American English combine all available articulatory mechanisms to produce a low third formant for /r/" (p. 163).

Gender Differences in the Constrictions for /r/

Gender differences in the acoustic characteristics of most speech sounds have been clearly documented. Those differences are typically described in terms of lower formant frequencies in males that simply reflect their overall longer vocal tracts. That said, as suggested in Chapter 1, the picture is likely more complicated. As Hagiwara (1995) reminds us, data and analysis by Fant (1963) indicated that male and female vocal tracts are actually organized (i.e., proportioned) differently. The average length of the adult male oral cavity (8.25 cm) is shorter than the average length of the corresponding male pharyngeal cavity (9.1 cm), yielding a ratio of

[1]Thankfully, it is not necessary to wait for those studies to be completed as ultrasound is not the only alternative approach available to remediate /r/ with evidence of effectiveness.

0.91. The average adult female oral and pharyngeal cavities are equal in length (7.0 cm each), yielding a ratio of 1.0. If the acoustic quality associated with /r/ is a function of the combination of the oral and pharyngeal constrictions, Hagiwara (1995) has suggested that males and females would need to create those constrictions in different places. In particular, he suggested that for females the oral constriction would need to be made relatively further forward in the mouth than for males.

TONGUE SHAPES FOR /r/

The idea of multiple constrictions that can be combined to generate the acoustic characteristics of /r/ implies that different combinations could be used successfully. In fact, prior to the use of ultrasound and continuing today, this has been implied by the ongoing discussion of different possible tongue shapes for an American English /r/. Phonetics textbooks (e.g., Shriberg et al., 2019) often suggest at least two such shapes: retroflex and bunched.

Might there be more than two tongue shapes for /r/? Delattre and Freeman (1968) used analysis of x-ray motion pictures from 46 adult speakers of British (n = 3) and American English (n = 43) to examine this issue. By grouping the tongue shapes they observed by visual similarity, they suggested there could be at least eight possible shapes (though only six of these were used by the speakers of American English). It is worth noting that although Delattre and Freeman included speakers from several different American dialect regions, they did not find any association between particular tongue shapes and particular dialects.

On a related note, the analysis confirmed the theoretical discussion about constrictions in both the oral and pharyngeal cavities for American English /r/. At the same time, the two shapes they associated with British English "only have ONE clear constriction, either at the pharynx . . . , or at the palate" (Delattre & Freeman, 1968, p. 42).

Another study of tongue shapes for /r/ was conducted by Westbury, Hashi, and Lindstrom (1998). In this case, they used the x-ray microbeam (see Chapter 1) to collect data from 53 American English adults, mostly from the Midwest. Tongue shapes were generated by connecting the positions of adjacent pellets using straight lines. They then calculated a set of three angles created among the four pellets and grouped them into similar shapes based on the sets of angles. Based on the pellet positions at the beginning of voicing for the /r/ in the word *row* and from the /r/ in *street*, the analysis resulted in four different tongue shapes.

Jakielski and Gildersleeve-Neumann (2018) mention that there may be three unique tongue shapes for American English /r/:

> A *bunched rhotic* [emphasis added] with the tongue tip down and the highest point of the tongue raised in a mid position toward the center of the mouth or the hard palate. The back edges of the tongue touch the back molars. The *alternative tongue tip rhotic* [emphasis added] has the tongue raised midway toward the back of the hard palate and the back edge of the tongue touching the back molars. The lips are rounded. However, some speakers of English produce [ɝ] using a *retroflex tongue* [emphasis added] gesture, that is, with the tongue tip curled up and back. (pp. 83–84)

Hagiwara (1995) presented data from 15 speakers and also suggested three tongue shapes for /r/ (which he termed *tip-down*, *blade up*, *tip up*) which appear to be quite similar to the three types just highlighted from Jakielski and Gildersleeve-Neumann.

Clearly this creates a somewhat confusing picture. How many different shapes are there for American English /r/? Two? Three? Four? Six? This is clearly not a trivial question. Clinicians need to know what target shape or shapes to aim for in therapy. Although some studies of these different versions of /r/ have suggested acoustic differences among them, not all investigators report differences. It is worth noting that the studies of the shapes were all based on productions that sounded like a perfectly acceptable /r/.

Although different shapes can be identified, researchers have suggested that not all the differences are clinically relevant. Westbury and colleagues (1998) noted that "(o)bserved differences in /ɹ/ are probably not significant at levels related to categorical perception (Is this /ɹ/ or /l/?), [or] clinical assessment (Is this a 'bad' /ɹ/ requiring therapy?)" (p. 221). Likewise, a 2016 paper by Mielke, Baker, and Archangeli discussed the eight different types proposed by Delattre and Freeman (1968) and suggested that these types "are exemplars of categories and not clearly useful as prototypes" (p. 103). In other words, they may be visually different, but the differences among them may not be meaningfully different for therapy. Even if they were, with the exception of the direct visual feedback of ultrasound, which is not likely to be available to most of our patients any time soon, we are not currently able to give specific enough feedback to allow them to produce very many different tongue shapes.

Giving instructions for making the distinction between the classic categories of *retroflex* and *bunched* would seem to be much more straightforward in terms of therapy targets. The former requires that the tongue tip be raised up above the midplane of the tongue body, while the latter suggests that the tongue tip remain clearly below that midplane. In addition, the emerging evidence from the treatment options to be discussed here also implies that the distinction into two basic types should work in most cases.

Box 2–1

Although many tongue shapes have been proposed for /r/, assuming that there are two tongue shapes for /r/ (bunched and retroflex) remains the most practical approach.

The two classic tongue shapes (retroflex and bunched) are illustrated in Figures 2–1 and 2–2, respectively.

Which Shape Is the Most Common?

As suggested in Chapter 1, anecdotal reports suggest that clinicians tend to focus heavily on a retroflex tongue shape. There are two possible explanations for this. First, a retroflex tongue shape may be more visible in the mouth and thus may be more easily modeled and imitated. This at least potentially increases the likelihood of treatment success.[2] The second possibility is that many clinicians may simply assume that the retroflex tongue is the most commonly used shape. If it is more common, this would also increase the likelihood of success with that shape. It would also mean that it would be less likely that we would have to switch to the other shape (which may be confusing for some patients) if our first choice was not helping them to produce a correct /r/. However, starting with a retroflex /r/ target makes less sense if a bunched tongue shape is actually more common.

No large-scale studies of the frequency of /r/ tongue shapes appear to have been conducted to date, and the available reports are somewhat contradictory. Secord and colleagues (2007) commented that "the retroflex tongue shapes appear to be less common than other types across the normal population of American English speakers" (p. 142). Mielke, Baker, and Archangeli (2016) compiled findings from several studies including most of those discussed earlier. They concluded that some speakers use a retroflex shape exclusively, while others use a bunched shape exclusively. A third group uses both shapes, but some of these did so somewhat randomly, while the remainder varied their shape specifically by phonetic context (i.e., bunched more often in vocalic and postvocalic contexts and retroflex in prevocalic contexts). Mielke and colleagues then added to the available evidence using ultrasound images from 27 adult American English speakers producing /r/ in a variety of phonetic contexts. They

[2]It has been suggested to me that this visibility assists with establishing the sound in isolation but that some children (once /r/ is well established) spontaneously switch to a more bunched tongue shape as therapy progresses.

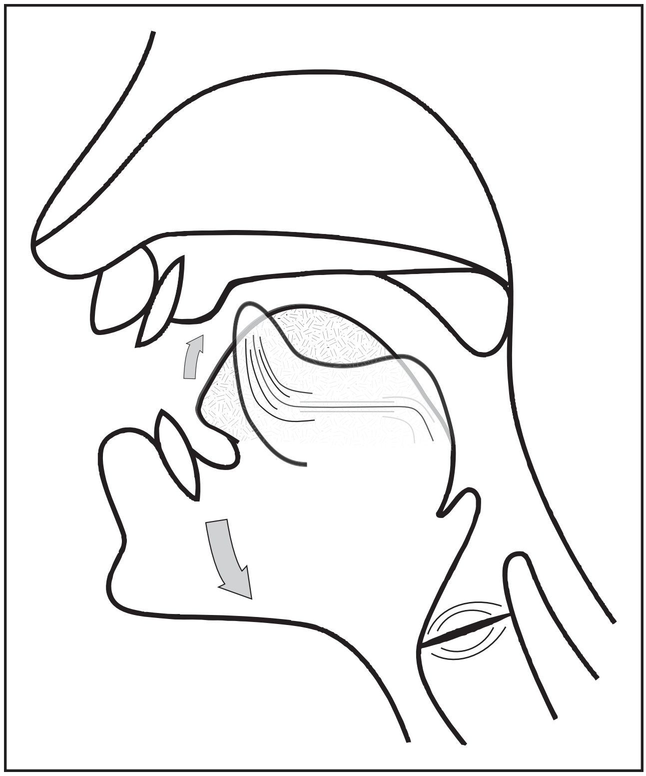

Figure 2–1. Retroflex tongue shape for /r/. From *Phonetic Science for Clinical Practice* by Kathy J. Jakielski and Christina E. Gildersleeve-Neumann, 2018, p. 85.

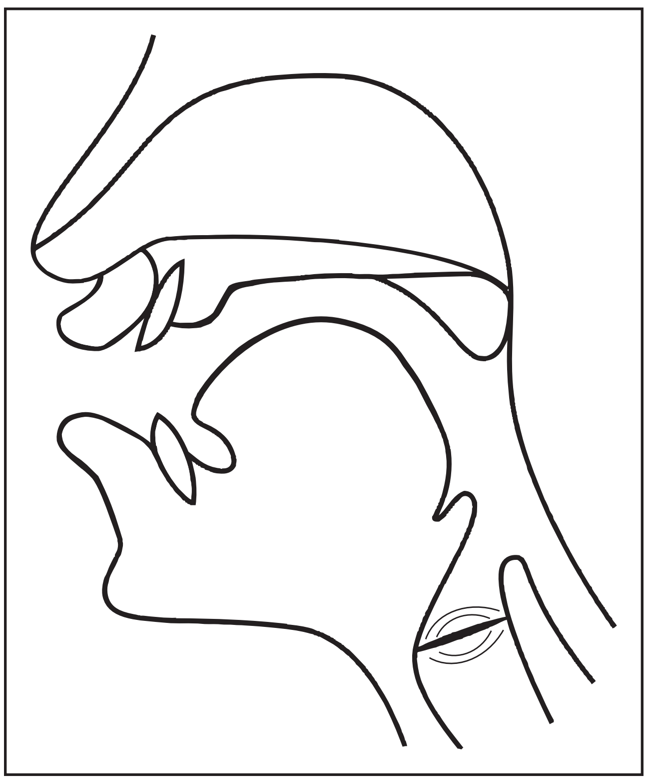

Figure 2–2. Bunched tongue shape for /r/. From *Phonetic Science for Clinical Practice* by Kathy J. Jakielski and Christina E. Gildersleeve-Neumann, 2018, p. 84.

found that 16 of the 27 participants used only a bunched /r/, 2 used only a retroflex /r/, and the remaining 9 used both. These data would suggest that a bunched shape might be more common. Another review by Boyce, Tiede, Espy-Wilson, and Groves-Wright (2015) was conducted to examine whether dialect might influence the choice of tongue shape for /r/. They concluded that "the simplest explanation of the data reviewed here is that the choice of tongue shape for /ɹ/ is a matter of individual variation rather than regional dialect" (p. 4).

Together the available evidence does not allow us to say with any certainty that a single tongue shape is used by most speakers. Thus, it is not clear which tongue shape is going to be optimal for most patients in therapy. At the very least, clinicians need to be prepared to change the target shape if progress is limited. The need to be flexible about the choice of tongue shape to be targeted is supported by findings from an ultrasound study by McAllister Byun, Hitchcock, and Swartz (2014). These authors stated that for both ultrasound and other treatment approaches:

> It is not optimal to target a single tongue shape for all clients; instead clients should be offered opportunities to explore different tongue shapes to find the configuration that is most facilitative of perceptually accurate rhotic sounds. (p. 2128)

Box 2–2

It is not clear whether a bunched or retroflex tongue shape is more common.

Some speakers use only one version exclusively.

Other speakers vary their shape depending on the phonetic context.

TONGUE BRACING FOR /r/

Another challenge for /r/, which was mentioned in Chapter 1, is limited tactile feedback. As a liquid consonant, production of /r/ only involves a relative narrowing of the vocal tract. As such, unlike many other consonant sounds, the tongue appears to have limited contact with the rest of the vocal tract, and the amount of tactile feedback available would appear to be limited.

Limited contact does *not*, however, mean zero contact or zero tactile feedback. Some tactile feedback may actually be available. The assumption of no contact likely comes from our reliance on two-dimensional images such as those from the tracing of x-ray images; even ultrasound only gives us a two-dimensional view at any one time.[3] Thus, we seem to have been ignoring most of the rest of the tongue. It has been suggested for some time (e.g., Stone, 1990) that speakers may brace the back or sides of the tongue for much of speech. In 2013, Gick and his colleagues reviewed available electropalatography (EPG), electromagnetic articulography (EMA), and ultrasound data and concluded that this was true; speakers brace their tongues against the palate and/or jaw for almost all speech sounds. The amount of bracing likely varies by speech sound. Specific to /r/, Bacsfalvi (2010) reported that (based on coronal view ultrasound images) many speakers appear to exhibit bracing of the posterior part of the tongue against the upper molars that results in a midline groove in the tongue.

Why might this be important for /r/? Given different tongue shapes for /r/, the amount or type of bracing (and therefore the nature of any tactile feedback generated) *may* vary across those shapes. Figure 2–3 shows an EPG image and x-ray tracing for a bunched /r/. In this case, bracing is likely against the upper teeth and palate; but, bracing for a retroflex /r/ (a tongue body low within the oral cavity and minimal contact with the palate) might be against the lower teeth or jaw. An EPG image in such cases might not be very helpful as no contact would be visible.[4] The choice of which tongue shape to target in therapy would therefore mean potentially different kinds of available tactile feedback. This may account for why changing the tongue shape target may lead to success for some patients; perhaps it is at least partially because of altered tactile feedback.

One final comment relative to tongue bracing for /r/. This should not be interpreted as a total blockage of the airstream. It only anchors the tongue at the sides. Coronal (face on) views with ultrasound have shown that during production of /r/ (particularly a bunched /r/), the center of the tongue is pulled down slightly from the sides forming a central channel. Neal (2020) referred to this as a "U-shaped tongue" for /r/.

MORE THAN ONE KIND OF /r/?

Setting aside the tongue shape question, it has been suggested that one of the reasons that /r/ is both difficult to learn and difficult to remediate is that it is not just one sound. It may vary in the way it is produced

[3]Rotating the ultrasound probe allows for a left-to-right view of the tongue to complement the typical front-to-back view.

[4]I must confess that prior to discovering this work on tongue bracing, and prior to learning that bunched /r/ may be at least as frequent as retroflex /r/, I was convinced that EPG therapy would be pointless for /r/ because there would be limited visual feedback. As shown in Chapter 9, that was an incorrect assumption on my part.

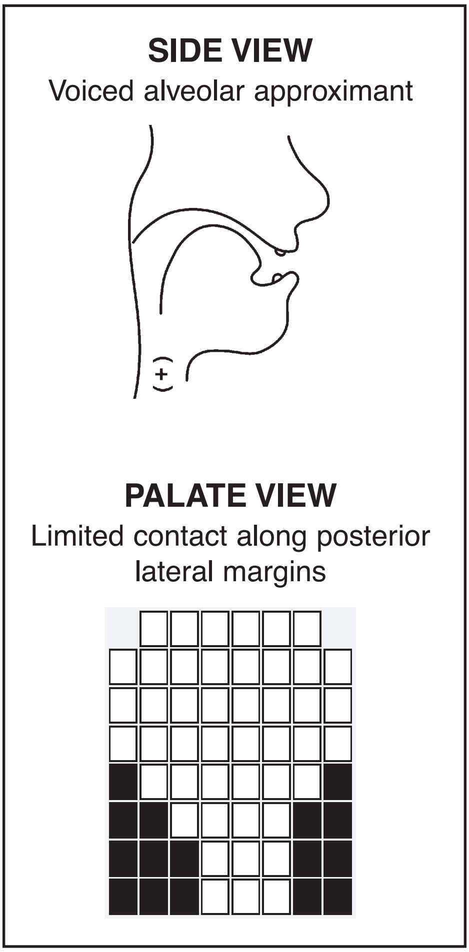

Figure 2–3. Tongue shape for /r/ (*upper image*) and electropalatography contact pattern (*lower image*) for a bunched /r/. From *Seeing Speech: A Quick Guide to Speech Sounds* by Sharynne McLeod and Sadanand Singh, 2009.

depending on the phonetic context in which it is produced. The fact that some speakers appear to use different tongue shapes in different contexts supports this. What would logically follow is that for some speakers it may be necessary to treat each phonetic context for /r/ individually.

The possibility of different versions of /r/ was implied in Chapter 1 during the discussion of the different phonetic symbols to be used for /r/. However, the different symbols being used for this sound only reflect differences in the linguistic function of /r/ within the syllable (consonant versus vowel), and its position in the word (pre- versus postvocalic for the consonant form and stressed versus unstressed syllable for the vocalic form). Acoustically, consonant and vocalic /r/ are the same (they differ primarily on duration and loudness).

Proponents of treating many different kinds of /r/ have suggested a much finer set of divisions that are based on the specific phonetic contexts in which /r/ might occur. One example of such a set of divisions (adapted from Ristuccia, 2002) is shown in Table 2–1.

The argument for different types of /r/ appears to be based on the idea that the surrounding contexts can influence how it is produced. There is little dispute about this. Context matters for /r/ as it does for every other speech sound. For example, consider /s/. When produced before a rounded vowel, speakers often round their lips during the production even though lip rounding is not usually associated with /s/. However, when /s/ is produced before an unrounded vowel, there is typically no lip rounding. Likewise, in faster, more casual speech, front or back vowels tend to be produced closer to the center of the vowel space than if they were being produced by themselves or in slower, careful speech (also called *clear* speech). All of these adjustments are examples of what is called *coarticulation,* which represents our very natural tendency to make the process of speaking more efficient. The ideal movement patterns for sounds spoken in isolation are modified to take into account where the articulators have come from and/or where they will be going next. Put another way, we take shortcuts when we produce a sound in the context of other sounds. We do so because this allows us to produce speech more quickly and with less

Table 2–1. Different Types of /r/

Context	Example Word	Context	Example Word
/ɝ/ initial	earth	/or/ medial	cork
/ɝ/ medial	fern	/or/ final	door
/ɝ/ final	fur	/ir/ initial	ear
Prevocalic (initial) /r/	red	/ir/ medial	zero
/ɚ/ medial	perform	/ir/ final	deer
/rl/ medial	Charlie	/ɛr/ initial	airplane
/rl/ final	girl	/ɛr/ medial	fairy
/ɑr/ initial	art	/ɛr/ final	hair
/ɑr/ medial	barn	/ɑɪr/ initial	Ireland
/ɑr/ final	far	/ɑɪr/ medial	fireplace
/or/ initial	organ	/ɑɪr/ final	tire

Note. Adapted from *The Entire World of R Instructional Workbook: A Phonemic Approach to /r/ Remediation* by C. Ristuccia, 2002. Say It Right. Copyright 2002 by Say It Right.

effort. We also likely do so because we can get away with it; there is little effect on our listeners (i.e., they usually do not notice), and communication is rarely affected. However, the idea that every phonetic context creates a unique sound form that must be learned (and by extension taught in therapy) only seems to have been applied to /r/.

Is /r/ a special case? Does its complexity (e.g., those three different constrictions to be coordinated) mean that the amount of motor learning required is simply greater than for other sounds? Do some speakers need to focus on each very specific context one at a time in order to master /r/? Maybe. The implication is that the mixing of contexts that we typically do in therapy may be too confusing for some children. Perhaps some of them need to work with a much narrower set of contexts to allow them to learn the sound. Two general observations support this notion. First, some children come to us with what at first appears to be inconsistent accuracy. In some of these cases, however, if we take the time to examine the specific contexts of /r/, we discover that they are consistently correct in some contexts and consistently incorrect in some other contexts. Such children might benefit from treating /r/ as more than one sound. Second, recall the previous discussion on the frequency of different tongue shapes for /r/. The fact that some speakers vary the tongue shape they use for /r/ depending on the phonetic context suggests that they are treating it as several sounds.

These observations, however, do not prove that such an approach would be effective in therapy. What is needed is direct empirical evidence showing that a narrow focus on specific contexts helps some children learn /r/. Clinicians using this approach report anecdotally that it appears to work, but there does not appear to have been any published studies of the use of the different kinds of /r/ listed in Table 2–1. Two older studies do provide some related findings. Both looked at whether a narrow focus on one particular form of /r/ might lead to generalization to other forms. In both cases, the three versions of /r/ discussed in Chapter 1 (/r/, /ɝ/, /ɚ/) were used. Elbert and McReynolds (1975) treated 12 children aged 6 to 11 years using traditional sound shaping therapy three times per week for 10 to 15 min. Findings indicated that "(m)ost subjects increased the number of correct responses to untrained items in several allophonic categories regardless of the specific allophone taught" (p. 386). Hoffman (1983) conducted a similar study with 12 children age 5;6 to 7;10 who received sound shaping therapy twice a week for 30 min. The outcomes were similar. All children generalized to untrained forms of /r/ regardless of which form was trained. Together these findings suggest that a narrow focus on either consonantal or vocalic /r/ may help *some* children master this sound. Perhaps limiting our treatment focus to one particular context may help some children.

As appealing as the prospect is of jumping in and working on each narrow phonetic context one at a time might be, a caveat is in order.

Although it may work for some patients, using such an approach may prove counterproductive. It may actually limit motor learning. In Chapter 5, principles of motor learning are discussed. One of these involves the idea that we should practice speech sound targets in a variety of contexts to encourage automaticity or flexibility. The ultimate goal for all of our patients is that they should be able to produce any speech sound in any valid phonetic context whenever they need to do so. Sticking too narrowly or for too long to specific contexts may actually discourage such flexibility; it risks therapy becoming the learning of particular motor patterns by rote (the exact opposite of flexibility). Mixing a variety of different contexts into our therapy needs to happen at some point to encourage flexibility in motor planning and motor execution.