Importance of the auditory perceptual target to the achievement of speech production accuracy

Importance de la cible perceptive auditive dans l'atteinte d'une production adéquate de la parole

Douglas M. Shiller
Susan Rvachew
Françoise Brosseau-Lapré

Abstract
The purpose of this paper is to discuss the clinical implications of a model of the segmental component of speech motor control called the DIVA model (Directions into Velocities of Articulators). The DIVA model is implemented on the assumption that the infant has perceptual knowledge of the auditory targets in place before learning accurate production of speech sounds and suggests that difficulties with speech perception would lead to imprecise speech and inaccurate articulation. We demonstrate through a literature review that children with speech delay, on average, have significant difficulty with perceptual knowledge of speech sounds that they misarticulate. We hypothesize, on the basis of the DIVA model, that a child with speech delay who has good perceptual knowledge of a phonological target will learn to make the appropriate articulatory adjustments to achieve phonological goals. We support the hypothesis with two case studies. The first case study involved short-term learning in a laboratory task by a child with speech delay. Although the child misarticulated sibilants, he had good perceptual and articulatory knowledge of vowels. He demonstrated that he was fully capable of spontaneously adapting his articulatory patterns to compensate for altered feedback of his own speech output. The second case study involved longer-term learning during speech therapy. This francophone child received 6 weeks of intervention that was largely directed at improving her perceptual knowledge of /ʃ/, leading to significant improvements in her ability to produce this phoneme correctly, both during minimal pair activities in therapy and during post-treatment testing.

Abrégé
Le but de cet article est de décrire les implications cliniques d'un modèle de la composante segmentale du contrôle moteur de la parole, plus précisément du modèle DIVA (« Directions into Velocities of Articulators »). Le modèle DIVA repose sur la prémisse que le nourrisson possède la connaissance perceptive des cibles auditives avant d'apprendre à produire correctement les sons, et suggère que les difficultés de perception de la parole engendrent une parole imprécise et une articulation inexacte. Nous démontrons à l'aide d'une revue de la littérature que les enfants présentant un trouble phonologique ont, en moyenne, des difficultés significatives avec la connaissance perceptive des sons qu'ils ne prononcent pas correctement. En se basant sur le modèle DIVA, nous posons l'hypothèse qu'un enfant qui présente un trouble phonologique et qui possède une bonne connaissance perceptive de la cible phonologique fera les ajustements articulatoires appropriés pour atteindre les cibles phonologiques. Nous présentons deux études de cas pour appuyer cette hypothèse. La première étude de cas implique un apprentissage à court terme dans une tâche en laboratoire par un enfant présentant un trouble phonologique. Malgré le fait que l'enfant n'articulait pas correctement les consonnes fricatives, il avait une bonne connaissance perceptive et articulatoire des voyelles. Il a démontré qu'il était pleinement capable d'adapter spontanément ses patrons articulatoires à de la rétroaction modifiée de sa propre parole. La deuxième étude de cas implique de l'apprentissage à plus long terme lors d'intervention en orthophonie. Cet enfant francophone a reçu six semaines d'intervention largement dirigée à améliorer la connaissance perceptive du phonème /ʃ/, menant à une amélioration significative de son habileté à produire ce phonème correctement lors d'activités de paires minimales en thérapie et lors de l'évaluation après la fin de l'intervention.

Keywords: speech sound disorders, speech motor control, speech perception, and speech therapy

Douglas M. Shiller, Ph.D., S-LP(C)
Université de Montréal
Montreal, Quebec
Canada

Susan Rvachew, Ph.D., S-LP(C)
McGill University
Montreal, Quebec
Canada

Francoise Brosseau-Lapré, M.Sc.A., S-LP(C)
McGill University
Montreal, Quebec
Canada

Accuracy

When designing an intervention for a child with primary speech delay (SD), the speech-language pathologist will typically begin with a description of 'what is wrong' in the child's overt speech. Over the past three decades there has been enormous change in the theoretical constructs used to conceptualize speech errors. Take for example, this excerpt from a child who was assessed at the age of 4;8 (Rvachew & Brosseau-Lapré, 2010): [bebi jeɪn dɑʊn ɪn gɑs̪ tʰæni ʌpʰ] (" baby laying down and dog's standing up"). Traditionally, one would describe omissions, distortions, and substitutions of segments, noting for example the child's substitution of [j] for /l/ in the word 'laying' (Van Riper, 1963). A phonological process analysis (Hodson, 2004) would take note of patterns of error in the child's speech such as the consistent reduction of the consonant sequences (e.g., /nd/, /gz/, /st/). Meanwhile, nonlinear phonological theories have focused our attention on interactions between features, segments and the prosodic aspects of the phonological system, allowing an explanation for the child's production of the word 'dog' that involves spreading of Dorsal from the place node of the coda to the place node of the onset segment combined with delinking of the coda itself from the skeletal tier (Bernhardt, Stemberger, & Major, 2006; Bernhardt & Stoel-Gammon, 1994). This historical shift in focus from the surface characteristics of the child's segment errors to a description of the child's underlying phonological knowledge[1] has led to the development of more efficient and effective approaches to speech therapy (Klein, 1996; Pamplona, Ysunza, & Espinoza, 1999). Nonetheless, the majority of children with SD make slow progress, failing to achieve normalized speech prior to school entry (Rvachew, Chiang, & Evans, 2007; Shriberg, Kwiatkowski, & Gruber, 1994), suggesting that these advances in phonological theory are not enough. Efficacious intervention programs require us to go beyond describing what the child is doing wrong and move toward explaining why the child is making those specific errors (Stackhouse & Wells, 1993). As highlighted by Bernhardt, Stemberger, and Charest (this issue), the task of imagining the possible sources of the child's errors requires that the speech-language pathologist consider models of language processing. In this paper, we begin with a discussion of a model of the segmental component of speech motor control. The DIVA model (Directions into Velocities of Articulators) is supported by research that ranges from computational modeling to clinical investigations involving behavioral and neuro-imaging methods (Callan, Kent, Guenther, & Vorperian, 2000; Ghosh, Tourville, & Guenther, 2008; Guenther, 1995; Guenther, Hampson, & Johnson, 1998; Perkell et al., 2000; Perkell et al., 1997).

The DIVA model contains a number of modules — each its own separate neural network — that capture the various steps in the transformation from the abstract phoneme string (the model's input) to the output articulatory motor sequence. The modules are connected by synaptic weights that implement the transformations, or *mappings*, between these representations. The model accounts for speech production development as the acquisition of three such mappings: the *phoneme-to-auditory* mapping, the *auditory-to-articulatory directional* mapping, and the *articulator-to-auditory* mapping. A critical assumption underlying the DIVA framework is that words are represented as a sequence of segments and that these segments are represented as spatio-temporal auditory goal regions (Perkell et al., 2000). While the model does not capture the full complexity of phonological representations (e.g., features at levels other than the individual segment, or the link between phonology and the lexicon) the implication is that the goal of the talker is to produce a specific auditory goal that will be perceived by the listener as the desired phoneme sequence, as opposed to producing a specific constellation of articulatory gestures. For example, if one wishes to convey the word 'we', comprising the phoneme sequence /wi/, one must produce an auditory product that corresponds to the phones [ui]. The corresponding auditory goals are invariant while not being point values. Rather, they are multi-dimensional regions in acoustic space as illustrated in Figure 1, depicting formant values appropriate for a child talker (represented here in only two dimensions, however, for the sake of simplicity). The talker could produce a variety of articulatory gestures that would result in the auditory goal of a second formant that is initially low and relatively close to the first formant but rising to a higher value that is much closer to the third formant in value (note that this characterization of the auditory goals in terms of relative locations of the formant frequencies allows for talker normalization so that the infant can learn to match his or her own speech output to auditory goals derived from adult input). The mapping between language-specific target phonemes and the corresponding auditory goal regions is learned very early in life but refined throughout childhood (Edwards, Fox, & Rogers, 2002; Hazan & Barrett, 2000; Kuhl, 2004; Maye, Werker, & Gerken, 2002; Nittrouer, 2002).

In the model, the auditory-to-articulatory directional and articulator-to-auditory mappings correspond to an internal model that is learned early in life on the basis of accurate sensory feedback. It is posited that the internal model is acquired during babbling as the infant learns to relate articulator movements to their orosensory and acoustic consequences. The critical role played by auditory input in the acquisition of auditory-motor mappings in the model is consistent with the empirical finding that hearing impairment in infancy delays the onset of the canonical babbling stage and reduces the amount and quality of speech-like babble produced by infants (Eilers & Oller, 1994; Koopmans-van Beinum, Clement, & van den Dikkenberg-Pot, 2001; Rvachew, Slawinski, Williams,

1 We use the term "knowledge" to refer broadly to information that is neurally encoded and accessible to the child, either with or without conscious awareness.

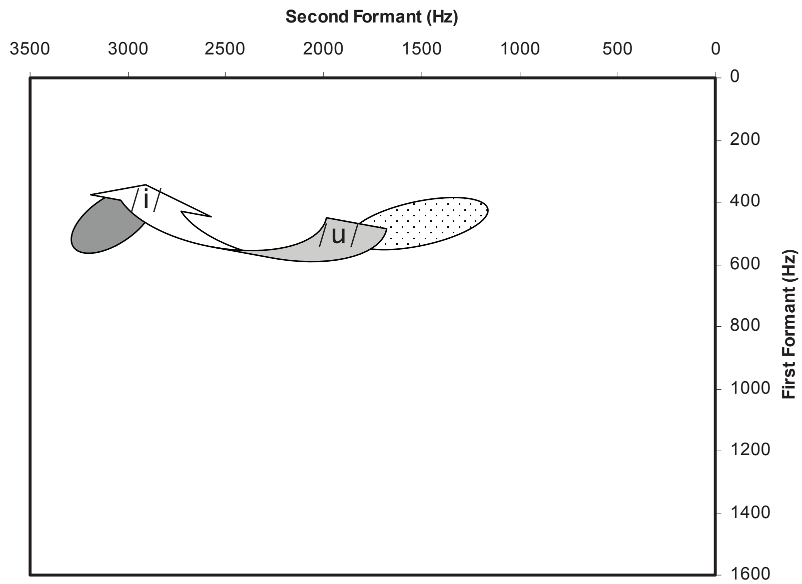

Figure 1: Illustration of hypothetical auditory goal regions when producing the word 'we'.

& Green, 1999). During the early word learning phase, the DIVA model is implemented on the assumption that the infant has perceptual knowledge[2] of the auditory targets already in place before learning accurate production of the requisite speech sounds (Guenther, 2003; Guenther, Ghosh, & Tourville, 2006). Perceptual knowledge of the target is essential if the infant is to use auditory feedback effectively to learn to plan articulatory movements that will result in the desired phone-sequences. The model further postulates that auditory feedback is used intermittently throughout childhood to reset the parameters of these mappings so that the child can cope with maturational changes in the size and shape of the vocal tract (Callan et al., 2000).

Perkell et al. (2000) argue that it is unlikely that auditory feedback is used by mature talkers to control articulatory movements in real time because of relatively long neural transmission times. Rather, the internal model allows the talker to rapidly translate information about the current vocal tract configuration into an estimate of auditory feedback by way of the articulatory-to-auditory mapping, which can then be used to drive the system toward the auditory goal. The internal model further allows the talker to plan a trajectory of movement from the current auditory region (e.g., the /u/ location marked on Figure 1) toward the target auditory goal region (e.g., the /i/ location), using an articulatory trajectory that maximizes *economy of effort*. The planned articulation trajectory in turn leads to the planning and execution of specific muscle activation patterns.

The planning of the articulatory trajectory in auditory space also allows the model to achieve similar acoustic outcomes (e.g., similar formant values) using different articulatory configurations (*motor equivalence)*, as is commonly observed in speech. For example, the articulatory trajectories from rounded vowels or consonants into /i/ will vary predictably from those produced in other phonetic contexts (Nittrouer, Studdert-Kennedy, & McGowan, 1989). The production of /u/ probably requires a change in constriction location from the palatal to the velar area during the infant period as the vocal tract is reshaped (Ménard, Schwartz, & Boë, 2002, 2004). This aspect of the model has important clinical implications and may help to explain in part the superior results of phonological interventions over traditional approaches even for children with structural deficits (Pamplona et al., 1999). In short, it implies that the focus of intervention should be on the successful achievement of phonological goals rather than specific articulatory gestures.

Turning to potential explanations for primary speech delay (of unknown origin), we now turn our attention away from causes related to the execution of the motor action plans and focus on the factors that might disrupt the development of the internal model, i.e., the three associated mappings: the *phoneme-to-auditory* mapping, *auditory-to-articulatory directional* mapping, and *articulator-to-auditory* mapping. It is clear that the ability to process acoustic-phonetic information is central to all three mappings. The acquisition of the *phoneme-to-auditory* mapping (that constitutes the auditory target regions for phonemes) requires the infant to detect statistical regularities in speech input in order to identify language specific phone-categories and the acoustic goal regions that are associated with those categories. Auditory feedback during babbling is essential if the infant is to learn to predict the articulatory patterns that give rise to a given acoustic pattern (auditory-to-articulatory directional mapping) and to predict the acoustic outcome of specific vocal tract configurations (articulatory-to-auditory mappings). Although speech delay is a heterogeneous diagnosis and there are other potential explanations (Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997), there is evidence to support the hypothesis that a very large subgroup of children with speech delay has a primary problem with the processing of the acoustic-phonetic characteristics of the speech input (Rvachew, 2007).

It is now fairly well established that, on average, children with speech delay of unknown origin have significant difficulty with speech perception. We searched the titles of journal articles published by the American Speech-Language and Hearing Association and identified 14 papers published since 1952 in which the speech perception abilities of children with speech delay were compared with

2 Perceptual knowledge includes a language-specific strategy for deriving phonological structure from acoustic input at multiple levels of the phonological hierarchy (e.g., features, segments, syllables, words). Speech perception is also influenced by sensory and nonsensory factors such as attention.

the speech perception abilities of children with normally developing speech (studies involving childhood apraxia of speech or secondary speech delay were omitted). These studies described collectively the perceptual abilities of 325 children with speech sound disorders, aged 3 to 9 years. With one ambiguous exception[3], every study demonstrated unequivocal evidence for significantly poorer speech perception abilities on the part of children with delayed speech. Speech perception deficits were observed when the children listened to live-voice or recorded natural speech (Cohen & Diehl, 1963; Hoffman, Stager, & Daniloff, 1983; Kronvall & Diehl, 1952; Marquardt & Saxman, 1972; Rvachew, Ohberg, Grawburg, & Heyding, 2003; Sherman & Geith, 1967; Smit & Bernthal, 1983), digitally altered natural speech (Edwards et al., 2002; Monnin & Huntington, 1974; Raaymakers & Crul, 1988), and synthesized speech (Broen, Strange, Doyle, & Heller, 1983; Hoffman, Daniloff, Bengoa, & Schuckers, 1985; Rvachew & Jamieson, 1989). In these studies the children with speech delay were shown to have difficulties with both discrimination and identification tasks involving the perception of their own speech as well as speech produced by other talkers. Cohen's *d* (with Hedge's correction), calculated for the comparison of clinical and typical groups, ranged from $d = 1.35$ (Rvachew & Jamieson, 1989, Study 1, n = 12) to $d = 8.75$ (Kronvall & Diehl, 1952, n = 30), indicating very large effect sizes for each of the 13 studies that provided usable data.

The nature of the perceptual deficits observed appeared to be in the realm of acoustic-phonetic representations (i.e., knowledge of the specific acoustic cues that permit identification of the relevant phonetic categories). In many studies (e.g., Rvachew & Jamieson, 1989) the children were able to perform the task with live-voice stimuli, indicating phonological knowledge of the target contrast, even though they had significant difficulty with the experimental task that did not provide visual and other nonstandard cues to the test contrast. Munson, Edwards, and Beckman (2005) addressed this question by asking children to repeat nonwords in which word length was held constant but the phonotactic probability of phoneme sequences within the words was varied, yielding better repetition performance for high probability sequences than low probability sequences. As would be expected, absolute accuracy of repetition varied between typical and clinical groups and was significantly correlated with speech perception and speech production skills. However, children with SD did not show a greater disadvantage when repeating low-frequency sequences than did typically developing children, relative to repetition accuracy for high-frequency sequences. These findings suggest that difficulties with abstract phonological knowledge are not the source of the articulation errors that are observed in children with SD. Rather, these children have difficulties constructing word representations in the more primary perceptual domain, an interpretation that is reinforced by more recent investigations involving a long-term repetition priming paradigm (Munson, Baylis, Krause, & Yim, 2006). Specifically, when repeating nonwords, children with SD did not benefit from prior hearing of the nonwords during a passive listening task, indicating that they had difficulty forming new perceptual representations after brief exposure to the novel words; in contrast, their typically developing peers were able to store memory traces for new words after minimal exposures during passive listening that supported improved repetition accuracy on subsequent trials.

In short, the children's performance in these studies suggests that they have some phonological knowledge of the target contrasts but they differentiate the contrasting phonemes on the basis of nonstandard and unreliable acoustic cues leading to inappropriate auditory goal regions for each phoneme. Edwards, Fourakis, Beckman and Fox (1999) demonstrated the close relationship between perceptual deficits and speech production errors in a study of six children with speech delay. As in other studies, the children with speech delay were able to identify words such as 'cape' and 'cake' in a picture pointing task when the words were presented live-voice. Compared with children with normally developing speech, they had significant difficulty with the task when small portions at the ends of the recorded words were excised or when the amplitude of the vowel portion of the words was attenuated. The authors concluded that the children's perceptual representations for these words were "vulnerable to diminished redundancy in the acoustic signal. (p. 182)" Acoustic analysis of the children's productions of words such as 'Timmy' and 'kitty' suggested poor speech motor control, even though perceptually correct /t/ versus /k/ contrasts were produced by most of the children. Compared with the control group, the children with speech delay demonstrated poor control over speaking rate, greater overlap in the skewness and centroid values for intended /t/ and /k/ productions, and larger transition slope values from lingual consonants into vowels. The authors concluded that the children with speech delay "were less able to maneuver jaw and tongue body separately."

In summary, the DIVA model of speech motor control suggests that difficulties with the acquisition of the phoneme-to-auditory mapping during early childhood leads to imprecise speech and inaccurate articulation, since that mapping defines a principal goal of the speech motor system. In the model, a precisely defined auditory goal region forms the basis for the feedback of error signals that tune the feed-forward command for production of the

3 Sommers, Cox and West (1972) published the only study of the 14 located that did not show evidence of speech perception deficits for children with SD. In this study, 8 groups of 7 children were selected on the basis of grade (kindergarten or grade 1), speech status (articulation normal or defective) and stimulability (high or low scores on a stimulability test). They concluded that "Superior articulators had significantly better scores than the deviant and defectives on the oral sensory discrimination task, but scores on the auditory tasks were not significantly different. Comparison of the performances of /s/ and /r/ defectives revealed the latter group to be inferior on some auditory tasks compared with the superior articulators. (p. 579)" However, the published paper includes the data for the oral sensory discrimination task for all groups but omits the speech perception data for the groups with normal speech development and thus it was impossible to confirm the findings or calculate effect sizes from the data as reported.

sound. Therefore, only once the perceptual target is known will the child be able to learn the precise articulatory gestures required to produce the phoneme[4]. When the perceptual target is unknown (e.g., when a child identifies [w] as /ɹ/), the child will be unable to learn the articulatory gestures associated with the target /ɹ/ or will be unable to achieve carry-over of /ɹ/ production to spontaneous speech. When the child's perceptual category for a given phoneme is too broad and/or defined by inappropriate cues (e.g., when the child focuses on the second formant of word-final /ɹ/ rather than the third formant, as in /w/), the child's production of the target may be marked by distortions and/or inconsistent substitution errors.

We have argued that a large proportion of children with speech delay have difficulties with speech perception that will interfere with the acquisition of the phoneme-to-auditory mapping. We now turn to two case studies with the intention of further demonstrating the importance of knowledge of the auditory target to speech development. The first case study involves short-term learning in a laboratory task by a child with speech delay. In this case, the child was forced to adapt to altered auditory feedback during a speaking task. The case study demonstrates that, when the auditory target is known, at least some children with speech delay are capable of speech motor learning over a short time. The second case study involves the application of a speech perception approach to intervention over a six-week period with a French-speaking child with speech delay. This case study demonstrates that an intervention that focuses on improving auditory-perceptual knowledge of the therapy target can lead to improved articulatory accuracy.

Case 1

The central role played by auditory representations in speech production has been highlighted in a number of recent studies investigating the effect of altered sensory feedback on the control of speech movements (Baum & McFarland, 1997; Houde & Jordan, 1998; Jones & Munhall, 2000, 2003; McFarland & Baum, 1995; Nasir & Ostry, 2006; Purcell & Munhall, 2006a; Savariaux, Perrier, & Orliaquet, 1995; Shiller, Sato, Gracco, & Baum, 2009; Tremblay, Shiller, & Ostry, 2003; Villacorta, Perkell, & Guenther, 2007). In studies of sensorimotor adaptation (SA), sensory feedback during speech production is altered either by introducing a mechanical perturbation to the oral articulators (e.g., an intra-oral prosthesis that alters palatal shape), or through the use of real-time signal processing to directly manipulate acoustic spectral properties (e.g., fundamental frequency, or vowel formant frequencies). A central aim of these studies has been to investigate the extent to which talkers alter their control of articulator movements to reduce the impact of the perturbation on the achievement of acoustic outcomes. In other words, they are a direct test of the hypothesis that speech production is organized around the achievement of precise auditory targets.

While physical manipulations are an effective means of disrupting auditory feedback, their overall impact on speech production is somewhat complex due to their multi-sensory nature (tactile, proprioceptive and auditory) and the fact that they may reduce the available articulatory degrees-of-freedom (e.g., in the case of jaw fixation using a bite-block, or lip-fixation using a lip-tube). Using real-time signal processing, it is possible to more precisely manipulate properties of the speech acoustic signal without impacting other sensory modalities or interfering with articulator motion. Studies have used this approach to investigate sensorimotor adaptation in adult talkers to a range of acoustic manipulations, including fundamental frequency (Jones & Munhall, 2000, 2003), vowel formant frequency (Houde & Jordan, 1998, 2002; Purcell & Munhall, 2006a, 2006b; Villacorta et al., 2007), and fricative spectral properties (Shiller et al., 2009). These studies have all demonstrated that following a period of speech practice under feedback-altered conditions, talkers tend to adjust their speech output in order to reduce the perceived magnitude of the manipulation (i.e., compensation was observed). Importantly, these studies have also demonstrated a continued effect on speech output following the unexpected removal of the feedback manipulation, indicating that the change was not simply the result of direct feedback-based adjustments, but rather a change in the way articulator movements were planned in advance (i.e., motor learning, or adaptation).

The fact that adult talkers readily adjust their speech motor output in order to maintain (relatively) consistent acoustic outcomes provides strong, direct evidence for the primacy of auditory sensory goals in speech production (as opposed to goals defined in terms of specific vocal tract configurations, for example). While some questions remain as to the precise sensorimotor processes underlying SA, the phenomenon is consistent with models such as DIVA, in which ongoing comparisons between auditory feedback and desired auditory sensory outcomes are used to maintain the accuracy of internal models involved in speech motor planning. Indeed, Villacorta et al. (2007) recently demonstrated the ability of the DIVA model to capture numerous aspects of sensorimotor adaptation to an auditory feedback manipulation.

Given the success of the SA paradigm in demonstrating a central role for auditory targets in the speech production of healthy adults, we were interested in the possibility that it might similarly allow us to demonstrate a role for precise auditory goals in children with speech delays. If

4 While in the model, accurate speech perception and auditory feedback allow for the establishment of auditory target regions for different speech sound categories, it is presumed that with practice (i.e., repeated production attempts), a set of analogous somatosensory target regions are also learned. Somatosensory feedback is then used alongside auditory feedback in order to detect errors and maintain speaking accuracy. The inclusion of a somatosensory feedback subsystem and somatosensory goals provides the model with *part* of what would be necessary to acquire speech production skill in the absence of auditory input. However, the model relies upon an intact auditory speech perceptual system to establish those targets by informing the system, during the early "babbling" stage, about whether a given movement attempt has resulted in the production of a particular speech sound.

we are to consider the possibility that an impairment in auditory perceptual representations is a factor in speech delay, it is necessary to demonstrate that these children in fact strive to achieve precise acoustic goals. Otherwise, the status of auditory representations might simply not be expected to have a large impact, and therapy focusing on phonemic perception would not be expected to have much impact on speech production. To this end, we present a case study of a child (CH) with a primary speech delay (primarily impacting his production of sibilant fricatives) who underwent a test of sensorimotor adaptation to altered auditory feedback.

It is important to note that the goal of examining sensorimotor adaptation in a child with speech delay was *not* to directly evaluate the relationship between perception and production of his misarticulated phonemes. Indeed, targeting the child's misarticulated consonants would likely yield results that are difficult to interpret, as any number of factors — including deficits in sensory, motor or cognitive processes — could lead to a failure to adapt, thus providing little information about the child's speech motor control processes. Rather, the goal was simply to demonstrate that: 1) children with speech delay spontaneously use auditory feedback in order to maintain the accuracy of speech motor planning, and 2) children with speech delay strive to achieve precision with respect to their achievement of acoustic outcomes, rather than striving to achieve a specific sequence of articulatory movements. Such findings have been demonstrated in prior studies of sensorimotor adaptation in healthy adults, but never before in children with atypical speech development. To this end, the test of sensorimotor adaptation that was carried out in this child examined his production of a *previously mastered* phoneme: the vowel ɛ/ (as in "head").

Participant

CH is a 6;6 year-old native English-speaking boy with a speech sound disorder but no reported history of language impairment, and no history of hearing impairment. At the time of testing, CH passed a pure-tone hearing screening and an oral mechanism exam that revealed no structural or functional abnormalities of the articulators and surrounding structures. Age appropriate expressive language skills were confirmed using the Formulated Sentences subtest of the Clinical Evaluation of Language Fundamentals, Fourth Edition (Semel et al., 2003; standard score = 13, 84th percentile). Receptive language and non-verbal cognitive abilities were also confirmed to be age-appropriate using the Kaufman Brief Intelligence Test, Second Edition (Kaufman & Kaufman, 2004; Verbal standard score = 103, 58th percentile; Non-verbal standard score = 121, 92nd percentile). CH's diagnosis of speech delay was confirmed by a standard score of 68 (6th percentile) on the Goldman Fristoe Test of Articulation, Second Edition (Goldman & Fristoe, 2000). CH's speech errors included a substitution of [θ] for /s/ and /ʃ/, substitution of [ð] for /z/ and /ʒ/, substitution of [w] for /r/, and a substitution of [f] for /θ/.

Method

Sensorimotor Adaptation Task

Similar to a number of previous studies of SA using auditory feedback manipulations (Purcell & Munhall, 2006a, 2006b; Villacorta et al., 2007), the present manipulation involved a real-time shift in the frequency of the first formant (F1) during repeated productions of /ɛ/ within the target word "head". F1 frequency was increased by approximately 175 Hz, which had the effect of reducing the separation between F1 and F2, yielding a vowel that was perceptually closer to /æ/ ("had").

While seated in a sound attenuating testing room, CH was instructed to produce the word "head" five times at a comfortable rate and volume, after which he would pause for 4 seconds while a visual "reward" was presented on a nearby computer monitor. This sequence was repeated 52 times, for a total of 260 productions of the target word.

The speech acoustic signal was transduced using a head-worn microphone, amplified, processed (see below for details), and then presented back to CH through circumaural headphones. The experimental protocol included four phases, carried out in the following sequence: 1) 50 repetitions of the target word under conditions of unaltered feedback (baseline phase), 2) F1 shift introduced in 30 Hz steps over a period of 60 trials (10 repetitions per step; ramp phase), 3) 120 repetitions of the target word under conditions of maximum F1 shift (hold phase), 4) 30 productions following the sudden removal of the F1 manipulation (after-effect phase) to evaluate the persistence of any compensatory change in vowel output.

The auditory feedback manipulation was achieved using a commercial digital signal processor (VoiceOne, TC Helicon) that is designed to manipulate speech acoustic signals. The VoiceOne is capable of real-time source-filter modeling of the incoming vocalized acoustic signal, and hence is capable of altering the shape of the spectrum with minimal impact on fundamental frequency and harmonics. In the present study, the formant shift was restricted to the F1 range using a low-pass filter to apply the spectral shift only to the low-frequency (< 1000 Hz) portion of the signal.

Compensatory changes in /ɛ/ production were evaluated on the basis of digitized acoustic recordings of the subject's speech output. The acoustic signal was initially digitized at 44.1 kHz and subsequently low-pass filtered and down-sampled to 10kHz for the purpose of formant analysis. For each production of the target word "head", a 30 millisecond portion of the signal located at the vowel midpoint was subjected to a formant analysis utilizing the Burg algorithm within Praat (Boersma & Weenik, 2009). The analysis provided estimates of the first four formant frequencies, of which only F1 and F2 were retained for further analysis.

While the manipulation of vowel feedback involved an increase in F1 frequency, the corresponding perceptual change was likely related to a decrease in the difference between F1 and F2 (F2-F1), an acoustic measure that has

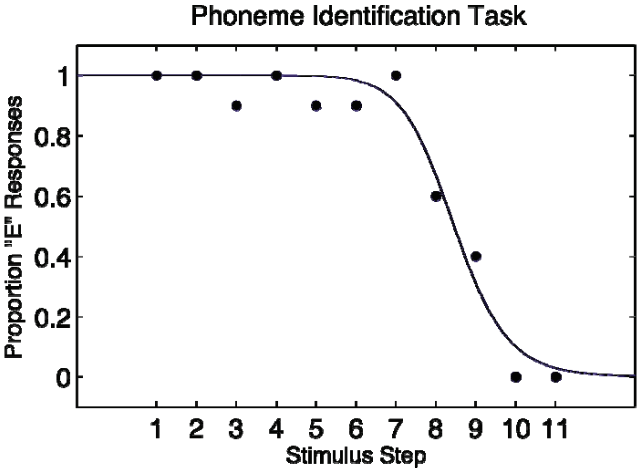

Figure 2: Response data for the vowel identification task. The filled circles show the proportion of /ɛ/ responses at each stimulus step. The solid line shows the best-fit logistic function.

Table 1
Mean change in F1 and F2 frequency (in Hz) relative to baseline at each phase of the sensorimotor adaptation task

Phase	F1 Change		F2 Change	
	Mean	SE[a]	Mean	SE[a]
Begin Hold	-35	3	-36	10
End Hold	-15	4	+132	6
Early After-Effect	-45	5	+48	9
Late After-Effect	-26	6	-4	8

[a] Standard error of the mean.

been found to be a stronger cue to tongue-height contrasts (e.g., /ɛ/ vs. /æ/) than F1 frequency alone (Kingston, 1991; Syrdal, 1985; Syrdal & Gopal, 1986). As a result, the acoustic analysis of CH's speech output focused on the F2-F1 feature, rather than changes in F1 frequency alone.

Phoneme Identification Task

Prior to the test of sensorimotor adaptation, a procedure was carried out to evaluate CH's perception of the /ɛ-æ/ contrast. The procedure involved the presentation of a synthetic vowel continuum that varied in 11 steps from [ɛ] and [æ]. The continuum was constructed by increasing the F1 frequency of a naturally produced [ɛ] token (spoken by an adult male in the context of the word "head") from approximately 550 Hz to 725 Hz (the talker's natural F1 frequency for "had"), using a signal processing approach similar to that used in the sensorimotor adaptation procedure. Following a practice run in which it was determined that CH understood the task and was able to correctly identify the endpoint vowel stimuli as /ɛ/ and /æ/, CH was presented with 10 repetitions of each of the 11 stimuli (always within the "h_d" context) in a fully randomized sequence. Following each stimulus presentation, CH indicated whether he had perceived the sound "E" as in "head" or "A" as in "had" by pressing the appropriate key on a keypad. In order to maintain his attention to the task, a child-friendly image was presented on a computer display following each block of 5 consecutive responses.

CH's response data were analyzed by first computing the proportion of /ɛ/ responses for each stimulus step (1.0 = 100% "E" responses), and then fitting a logistic function to the resulting data points in order to quantify the location and slope of the perceptual boundary between phoneme categories.

Results

Perception of the /ɛ–æ/ contrast

CH's response data for the vowel identification task (proportion of /ɛ/ responses for each stimulus step) are presented in Figure 2, along with the best-fit logistic function. The results show a sudden perceptual shift from /ɛ/ to /æ/ in the vicinity of stimulus 8-9, demonstrating an ability to perceive the contrast between these two vowel categories.

Sensorimotor adaptation

Baseline F1 and F2 frequencies were estimated from the final 10 productions of "head" under normal feedback conditions (immediately prior to the onset of the ramp phase). Mean F1 and F2 frequency were 753 and 2392 Hz respectively. Subsequent changes in vowel formant frequencies were evaluated by computing mean F1 and F2 values within four different blocks of trials (10 trials per block): 1) at the beginning of the hold phase (early training), 2) at the end of the hold phase (late training), 3) immediately following removal of the feedback manipulation (early after-effect), 4) at the end of the after-effect phase (late after-effect). Mean formant values for each block are presented in Table 1, and changes in F2-F1 are presented graphically in Figure 3.

While CH showed little change in F2-F1immediately following the ramped onset of the feedback manipulation (-1 Hz change), a compensatory change (i.e., an increase in F2-F1) was observed at the end of the hold phase (+147 Hz change). The compensatory F2-F1 change was found to persist immediately following removal of the feedback manipulation (though at +93 Hz, it was smaller than the effect observed at the end of the hold phase), indicating that underlying motor plans for the production of the vowel had in fact been altered (i.e., adaptation). By the end of the after-effect phase (following 20 productions under conditions of unaltered auditory feedback), CH's F2-F1 values had returned close to baseline (+22 Hz).

The reliability of these F2-F1 effects was evaluated using a one-way, independent measures ANOVA, treating each block of trials as a random sample of scores (N=10). An overall main effect of trial block was found ($F[4,42]=5.16$, $p < 0.01$). Post-hoc comparisons using Tukey's method revealed reliable differences ($p < 0.05$) between baseline

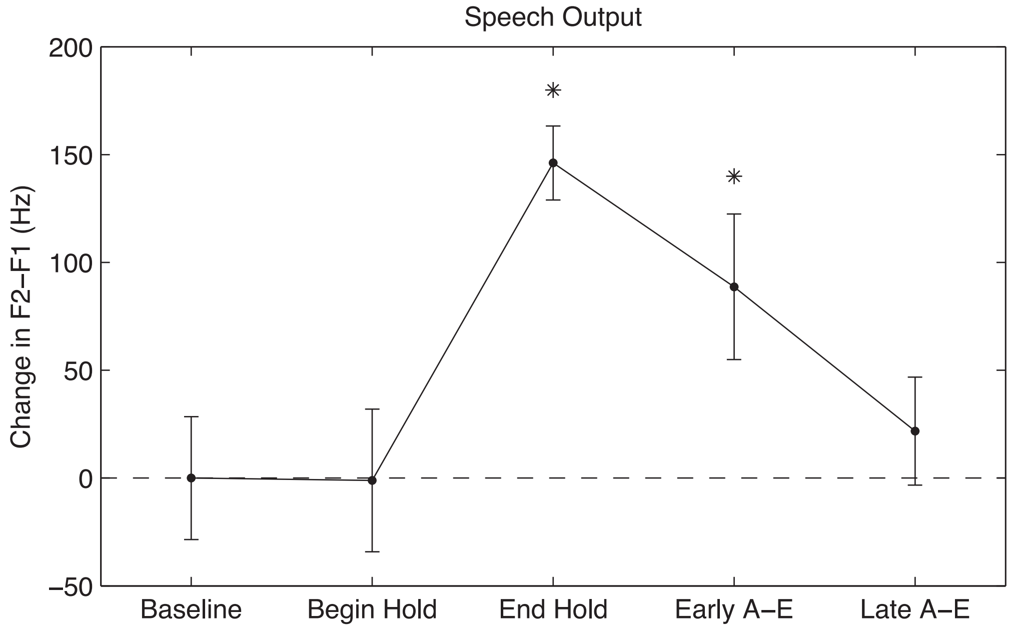

Figure 3: Mean change in F2-F1 (in Hz) at each phase of the sensorimotor adaptation task.

and late training blocks, between early training and late training blocks, and between the late training and late after-effect blocks (Figure 3).

Discussion

The finding that CH, a 6;6 year old child with a primary speech delay, readily adapted his speech output in response to a manipulation of auditory feedback adds a small but valuable piece of information to the present discussion about auditory perceptual goals in speech production. Phoneme identification testing indicated that the child had good perceptual knowledge of the phonological target. When auditory feedback was manipulated to create the impression that his speech output was deviating from this goal, he spontaneously adapted his speech output to the auditory feedback. His performance demonstrated accurate knowledge of the relationship between articulatory movements and relative formant frequency locations as well as sufficient speech motor control to achieve his speech goals. While only a single case, it nevertheless supports the notion that children with speech delays organize their control of speech production around the achievement of precise auditory goals (though, not necessarily accurate ones). It leads to the hypothesis that ensuring adult-like perceptual knowledge of phonological targets will facilitate the acquisition of articulatory knowledge of those targets. The next case study demonstrates that an intervention that targets perceptual knowledge of the phonological target can lead to improvements in articulatory accuracy.

Case 2

The Speech Assessment and Interactive Learning System (SAILS) is a computer game that was developed to teach children the appropriate acoustic-phonetic goal regions for commonly misarticulated phonemes (for literature review and video tutorial, see Rvachew & Brosseau-Lapré, 2010). The software presents children with recorded versions of single syllable words produced by adult and child talkers. The listener's task is to point to a picture of the word when they hear a well-produced version of the target and to point to an 'X" when they hear something that is not the target. The task was designed to reflect Locke's (1980) call for clinically relevant speech perception test procedures that assess the match between adult surface forms and the child's own internal representations for words, targeting those phonemes that the child misarticulates. This program has been shown to facilitate the acquisition of correct production of the target phoneme in a series of single subject experiments (Jamieson & Rvachew, 1992), a quasi-experiment (Rvachew, Rafaat, & Martin, 1999), and three randomized control trials (Rvachew, 1994; Rvachew, Nowak, & Cloutier, 2004; Wolfe, Presley, & Mesaris, 2003). For example, Rvachew (1994) was conducted with preschool children presenting with moderate speech sound disorders who received six therapy sessions once weekly, in all cases targeting /ʃ/ for which the children were unstimulable. All children received perceptual training in addition to traditional speech therapy but only a third of the children listened to various productions of the target /ʃ/, both articulated correctly and incorrectly, by completing the SAILS intervention modules targeting this phoneme. Children in the other conditions either listened to a single well-produced version of the word *shoe* contrasted with one version of the word *moo,* or to the words *cat* and *Pete.* In this study, the perception training component lasted for one third of each session while two-thirds of all therapy time was devoted to production training. The production training procedures were behaviorist in nature, involving phonetic placement, progressive approximation and practice with progressively longer utterances. Feedback was provided about the accuracy of the children's articulatory gestures and a high rate of accuracy was required before children could advance from one step of the treatment program to the next (for example, they were required to imitate syllables with 90% accuracy before practice with the imitation of words was introduced). Children who had completed the SAILS intervention modules made more progress with respect to production of /ʃ/ than other children in the study. In fact, children in the control group failed to achieve stimulability for /ʃ/ in isolation whereas children who received speech perception training learned to produce this phoneme in phrases.

Currently we are conducting a randomized control trial that involves a French version of SAILS (Essai Clinique Randomisé sur les Interventions Phonologiques). The francophone children that are enrolled in this trial receive 6 weeks of individual intervention directed at improving their articulation accuracy followed by six weeks of group intervention targeting phonological awareness skills. Half of the children in the trial are randomly assigned to receive individual therapy that is focused on improving their perceptual knowledge of their speech targets. The intervention differs from that employed in previous studies in that the proportion of time devoted to listening activities versus speech production practice is much greater. Furthermore, speech practice activities take place in the context of minimal pair activities that are designed to provide feedback about the communicative effectiveness of the child's speech. Phonetic placement and overt feedback about articulatory gestures are discouraged. Overall the program is designed to ensure that the child gains good perceptual knowledge of the target and then has opportunities to discover the articulatory movements that are necessary for accurate achievement of speech goals. In this case study we present the results for one child who is enrolled in this study. The child's performance will be described for the first 6-week period when she received individual therapy from a student speech-language pathologist under the supervision of the third author who is coordinating this trial.

Participant

Participant 1113 was four years eight months at the intake assessment and presented with a moderate speech delay. Her vocabulary skills were within normal limits, her score on the Échelle de vocabulaire en Images Peabody (EVIP; Dunn, Thérialut-Whalen, & Dunn, 1993) being at the 50th percentile rank. She also obtained a standard score of 103 on the matrices subtest of the Kaufman Brief Intelligence Test – Second edition (K-BIT-2; Kaufman & Kaufman, 2004) indicating average non-verbal IQ. Participant 1113 also passed the Oral Speech Mechanism Screening Examination - Third edition (OSMSE-3; St-Louis & Ruscello, 2000), revealing normal structure and function of the oral mechanism. At the present time, there is no normed and validated test of phonology available for French; clinicians typically use a language sample and their clinical judgment to qualify the degree of severity of the phonological impairment in this language. Participant 1113 obtained a diagnosis of a moderate speech delay by the community speech-language pathologist who had re-assessed her two weeks prior to her referral to the ECRIP research project. The Test Francophone de Phonologie (TFP) is currently being developed by Paul & Rvachew and contains 54 single words representing the characteristics of the phonology of Quebec French. On the TFP, administered during the intake assessment, participant 1113 did not produce responses spontaneously and therefore delayed imitation and immediate imitation were used in order to obtain responses for every test item. She obtained a percentage of consonants correct of 81 based on phonetic accuracy of each consonant articulation, i.e., omission, substitution, and distortion errors were scored as incorrect. Her error patterns included fronting of /ʃ/ to [s], reduced consonant clusters and deletion of syllables in multisyllabic words. Intelligibility in conversation was more affected and was severely reduced in unknown contexts.

Methods and Results

Pretreatment Assessment

Following the intake assessment, three specific therapy targets were selected for participant 1113, one of which was to improve auditory-perceptual knowledge of /ʃ/. Prior to the first therapy session, participant 1113 was asked to produce 20 words containing /ʃ/. Pictures of the target words were presented in four blocks of five items each, with the clinician naming each block before prompting the child to name the items in the same order. She obtained a score of 1 out of 20, producing [s] for all other items. During the same probe session, speech perception of /ʃ/ was assessed using the French version of the Speech Assessment and Interactive Learning System (SAILS, AVAAZ Innovations, Inc., 1994). Participant 1113 obtained a score of 50% on both the modules 'chat' [ʃɑ] *(cat)* and 'tache' [tɑʃ] *(spot)*, indicating poor perceptual knowledge of this phoneme.

Treatment

Intervention for participant 1113 consisted of three types of activities: SAILS, focused stimulation, and minimal pairs. SAILS is a computer game that uses a two-alternative forced choice identification task. The child listens to stimuli recorded from adults and children with and without speech sound disorders and needs to indicate whether each word presented is a good exemplar of the target or not. Each block contains five correctly and five incorrectly articulated target phonemes corresponding to typical misarticulations from younger children and children with SSD. During intervention, feedback is provided by the clinician when the child chooses the wrong response alternative and then the stimulus is repeated. The feedback includes a brief explanation as to why the presented stimulus did not match the child's response, and the child must then select the correct response to continue to the next trial. Participant 1113 completed a different SAILS module during each of the first three therapy sessions; approximately ten minutes were devoted to each module, which consisted of a practice block and two intervention blocks. In the module "chat", the practice block contrasts five adult productions of the word "chat" [ʃa] and five adult productions of the word [ma]; foil items in Block 1 are child productions of [ta] and [da]; foil items in Block 2 are child productions of [sa], [s̺a] and [ʃa]. The practice block of the "chaude" (hot) module contains adult productions of [ʃod] and adult productions of [mod]; foil items are child productions of [tod] and [dod] in Block 1 and child productions of [sod], [s̺od] and [ʃod]. In the "tache" module, which targets word-final /ʃ/, practice items are [taʃ] and [tap]; foil items are [tat] in Block 1 and [tas], [tas̺] and [taʃ] in Block 2. It should be noted that Participant 1113 enjoyed completing the SAILS modules.

Second, focused stimulation activities provided participant 1113 with many opportunities to hear words containing /ʃ/. For example, the clinician selected books that contained frequent repetitions of one or a few words containing the target phoneme. Activities involving toys were also used, for instance while playing with a farm the clinician repeated the words "cheval" [ʃəval] (horse), "cochon" [koʃõ] (pig) and "vache" [vaʃ] (cow) on numerous occasions (targeting /ʃ/ in all three word positions). Participant 1113 was never asked to produce the target words during these activities, but had opportunities to do so. If she attempted production of the target words the clinician would recast her attempted production if necessary by repeating her utterance and correctly producing the target word. No explicit feedback was given to participant 1113 regarding the accuracy of her productions. Focused stimulation activities were completed during the second, third and fourth therapy sessions, for five to seven minutes each.

Third, perceptual and production minimal pairs activities were used. During perceptual minimal pairs, participant 1113 had to identify whether the clinician produced the target word correctly or produced the child's mispronunciation. For instance, if the clinician produced the word "choux" [ʃu] (cabbage) properly, the child was expected to glue the picture of a cabbage in the garden but if the clinician said "sous" [su] (penny) the child was expected to take a penny placed on the table and to give it to the clinician. The clinician provided feedback to the child, represented the stimulus word and helped the child select the correct object if needed. During production minimal pairs activities, participant 1113 was required to produce the target word. Activities were designed so that she could not achieve her goals if she produced [s] instead of [ʃ]; for example if she said "ça" [sa] (this) instead of "chat" [ʃa] she could not obtain the cat stickers to complete the activity. Perceptual minimal pairs activities were carried out in the third, fourth and fifth therapy sessions. Performance was found to improve across the three sessions, with the child correctly identifying only 1/10 productions of [ʃ] in the first session, and 10/10 by the end of the third session. Production minimal pairs activities were completed during the fifth and sixth sessions. The child showed improvements between these two sessions, with word-initial [ʃ] improving from 0% to 100% correct, and word-final [ʃ] improving from 0% to 60% correct.

Post-treatment Assessment

Following the six therapy sessions, participant 1113 correctly produced 13 of the 20 probe words in delayed imitation; more specifically she correctly articulated all words containing /ʃ/ in the onset position (CV, CVC, CVCV and CVCVC word structures), 1 of 5 words containing /ʃ/ in the word-medial position which had been targeted during two therapy sessions and 2 of 5 target phonemes in the coda position (CVC word structure). The focus on speech perception during intervention probably allowed participant 1113 to develop an internalized perceptual-acoustic representation for /ʃ/ so that she was able to self-monitor and self-correct her own speech. The ultimate goal of the perceptual intervention is to allow the child to discover the articulatory gestures associated with the correct production of the phoneme so that she can accurately produce the target phoneme with greater frequency. The TFP was re-administered seven weeks later, following a six-week period of phonological awareness intervention. Participant 1113 produced all words in a delayed imitation task, and obtained a percent of consonants correct of 84.

General Discussion

According to the DIVA framework, the achievement of accurate speech production is wholly dependent upon a learned mapping between phonemes and the auditory goal regions that correspond to those phonemes (phoneme-to-auditory mapping). Knowledge of the auditory target allows the child to discover the predictive relationships between the various articulatory patterns that give rise to a given acoustic pattern (auditory-to-articulatory directional mapping) and the acoustic outcomes of specific vocal tract configurations (articulatory-to-auditory mappings). Knowledge of these relationships allows the talker to plan articulatory movements in order to achieve speech production goals with precision and economy of effort.

Studies that involve the manipulation of auditory feedback show that adult talkers readily adjust their speech motor behaviour in order to maintain (relatively) consistent acoustic outcomes, providing evidence for the primacy of auditory sensory goals in speech production. In the first case study, we demonstrated that a child with SD had an adult-like ability to adapt his speech motor output to achieve a phonological/auditory goal corresponding to the word 'head' (distant F2-F1 during /ɛ/) when feedback of his speech productions was manipulated to produce a percept similar to 'had' (i.e., close F2-F1 during /æ/). He was able to learn this task very quickly with no explicit instruction. The demonstration that this child strives to achieve precision with respect to his achievement of acoustic outcomes (rather than aiming to achieve a specific sequence of articulatory movements) suggests that his sibilant misarticulations may be due to mis-specified auditory goal regions for these phonemes rather than an inability to adjust articulatory patterns to achieve the necessary vocal tract configurations for accurate production. Without a direct test of his perceptual categorization of the sibilants, it is not possible to state this with absolute certainty, however.

In a second case study we demonstrated that a child who had very poor perceptual knowledge of /ʃ/ could make significant gains in articulatory accuracy for this speech sound with minimal speech production practice. This child received approximately 90 minutes of intervention for this phoneme but only about 15 minutes of this time, at the very end of six week intervention period, was devoted to overt speech production practice in the context of meaningful minimal pairs activities. Treatment activities involved primarily speech perception training, focused stimulation, and receptive minimal pair activities, designed to ensure that the child improved her perceptual knowledge of /ʃ/ as presented in a variety of syllable structures. When minimal pair production activities were introduced during the last two treatment sessions she quickly achieved success at these tasks. In pre- and post-treatment probes, she improved her performance from 5% to 65% correct, demonstrating correct articulation in a variety of word positions. Although she did not achieve consistently correct production of this phoneme, she demonstrated self-correction of her misarticulations very shortly after the introduction of speech practice activities.

Many studies conducted over the past five decades have shown that children with SD have significant difficulties with the perception of speech sound contrasts that they misarticulate. These children's speech perception difficulties may reflect auditory goal regions for phonemes that are overly broad, and hence overlapping with other phonemes. The DIVA model explains how these speech perception difficulties impact on the development of speech motor control and provides a rationale for the effectiveness of speech perception training as a means of facilitating children's response to speech therapy.

References

AVAAZ Innovations, I. (1994). *Speech assessment and interactive learning system* (Version 1.2). London, Ontario, Canada

Baum, S. R., & McFarland, D. H. (1997). The development of speech adaptation to an artificial palate. *J Acoust Soc Am, 102*(4), 2353-2359.

Bernhardt, B., Stemberger, J. P., & Major, E. M. (2006). General and nonlinear phonological intervention perspectives for a child with resistent phonological impairment. *Advances in Speech-Language Pathology, 8*, 190-206.

Bernhardt, B., & Stoel-Gammon, C. (1994). Nonlinear phonology: Introduction and clinical application. *Journal of Speech and Hearing Research, 37*, 123-143.

Boersma, P. & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.08) [Computer program]. Available from http://www.praat.org/

Broen, P. A., Strange, W., Doyle, S. S., & Heller, J. H. (1983). Perception and production of approximant consonants by normal and articulation-delayed preschool children. *Journal of Speech and Hearing Research, 26*(4), 601-608.

Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. *Journal of Speech, Language, and Hearing Research, 43*, 721-738.

Cohen, J. H., & Diehl, C. F. (1963). Relation of speech sound discrimination ability to articulation-type speech defects. *Journal of Speech and Hearing Disorders, 28*, 187-190.

Dunn, L., Thériault-Whalen, C., & Dunn, L. M. (1993). *Échelle de vocabulaire en images Peabody*. Toronto: Psycan.

Edwards, J., Fourakis, M., Beckman, M. E., & Fox, R. A. (1999). Characterizing knowledge deficits in phonological disorders. *Journal of Speech, Language, and Hearing Research, 42*, 169-186.

Edwards, J., Fox, R. A., & Rogers, C. L. (2002). Final consonant discrimination in children: Effects of phonological disorder, vocabulary size, and articulatory accuracy. *Journal of Speech, Language, and Hearing Research, 45*, 231–242.

Eilers, R., & Oller, D. (1994). Infant vocalizations and the early diagnosis of severe hearing impairment. *Journal of Pediatrics, 124*(2), 199-203.

Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2008). A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. *Journal of Speech, Language, and Hearing Research, 51*, 1183-1202.

Goldman R, Fristoe M 2000. Goldman Fristoe Test of Articulation. 2nd ed. Circle Pines, Minnesota: American Guidance Services.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review, 102*(3), 594-621.

Guenther, F. H. (2003). Neural control of speech movements. In A. M. a. N. Schiller (Ed.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Berlin: Mouton de Gruyter.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 96*, 280-301.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105*(4), 611-633.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *Journal of Phonetics, 28*, 377-396.

Hodson, B. (2004). Hodson Assessment of Phonological Patterns, 3rd Ed. Austin, TX: Pro-Ed.

Hoffman, P. R., Daniloff, R. G., Bengoa, D., & Schuckers, G. (1985). Misarticulating and normally articulating children's identification and discrimination of synthetic [r] and [w]. *Journal of Speech and Hearing Disorders, 50*, 46-53.

Hoffman, P. R., Stager, S., & Daniloff, R. G. (1983). Perception and production of misarticulated /r/. *Journal of Speech and Hearing Disorders, 48*(2), 210-215.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science, 279*(5354), 1213-1216.

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *J Speech Lang Hear Res, 45*(2), 295-310.

Jamieson, D. G., & Rvachew, S. (1992). Remediation of speech production errors with sound identification training. *Journal of Speech-Language Pathology and Audiology, 16*, 201-210.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: evidence from feedback perturbation. *J Acoust Soc Am, 108*(3 Pt 1), 1246-1251.

Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: the role of auditory feedback. *J Acoust Soc Am, 113*(1), 532-543.

Kaufman, A. S. & Kaufman, N. L. (2004) Kaufman Brief Intelligence Test, 2nd edition. AGS Publishing, Circle Pines, MN, USA.

Kingston, J. (1991). Integrating articulations in the perception of vowel height. *Phonetica, 48*, 129-179.

Klein, E. S. (1996). Phonological/Traditional Approaches to Articulation Therapy: A Retrospective Group Comparison. *Lang Speech Hear Serv Sch, 27*(4), 314-323.

Koopmans-van Beinum, F. J., Clement, C. J., & van den Dikkenberg-Pot, I. (2001). Babbling and the lack of auditory speech perception: a matter of coordination? *Developmental Science, 4*(1), 61-70.

Kronvall, E. L., & Diehl, C. F. (1952). The relationship of auditory discrimination to articulatory defects of children with no known organic impairment. *Journal of Speech and Hearing Disorders, 19*, 335-338.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews: Neuroscience, 5*, 831-843.

Locke, J. L. (1980). The inference of speech perception in the phonologically disordered child. Some clinically novel procedures, their use, some findings. *Journal of Speech and Hearing Disorders*, 45, 445-468.

Marquardt, T. P., & Saxman, J. H. (1972). Language comprehension and auditory discrimination in articulation deficient kindergarten children. *Journal of Speech ahd Hearing Research, 15*, 382-389.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, B101-B111.

McFarland, D. H., & Baum, S. R. (1995). Incomplete compensation to articulatory perturbation. *J Acoust Soc Am, 97*(3), 1865-1873.

Ménard, L., Schwartz, J., & Boë, L. (2002). Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *Journal of the Acoustical Society of America, 111*(4), 1892-1905.

Ménard, L., Schwartz, J., & Boë, L. (2004). Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized vowels from birth to adulthood. *Journal of Speech, Language, and Hearing Research, 47*, 1059-1080.

Monnin, L. M., & Huntington, D. A. (1974). Relationship of articulatory defects to speech-sound identification. *Journal of Speech and Hearing Research, 17*, 352-366.

Munson, B., Baylis, A., Krause, M., & Yim, D.-S. (2006). *Representation and access in phonological impairment.* Paper presented at the 10th Conference on Laboratory Phonology, Paris, France, June 30-July 2.

Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research, 48*, 61-78.

Nasir, S. M., & Ostry, D. J. (2006). Somatosensory precision in speech production. *Current Biology, 16*(19), 1918-1923.

Nittrouer, S. (2002). From ear to cortex: A perspective on what clinicians need to understand about speech perception and language processing. *Language, Speech, and Hearing Services in Schools, 33*, 237-252.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R. S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32*(1), 120-132.

Pamplona, M. C., Ysunza, A., & Espinoza, J. (1999). A comparative trial of two modalities of speech intervention for compensatory articulation in cleft palate children: phonological approach versus articulatory approach. *International Journal of Pediatric Otorhinolaryngology, 49*, 21-26.

Perkell, J., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., et al. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics, 28*, 233-272.

Perkell, J., Matthies, M., Lane, H., Guenther, F. H., Wilhelms-Tricarico, R., Wozniak, J., et al. (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication, 22*, 227-250.

Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *J Acoust Soc Am, 120*(2), 966-977.

Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *J Acoust Soc Am, 119*(4), 2288-2297.

Raaymakers, E. M., & Crul, T. A. M. (1988). Perception and production of the final /s-ts/ contrast in Dutch by misarticulating children. *Journal of Speech and Hearing Disorders, 53*, 262-270.

Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research, 37*, 347-357.

Rvachew, S. (2007). Phonological processing and reading in children with speech sound disorders. *American Journal of Speech-Language Pathology, 16*, 260-270.

Rvachew, S., & Brosseau-Lapré, F. (2010). Speech perception intervention. In S. M. L. Williams, & R. McCauley (Ed.), *Interventions For Speech Sound Disorders in Children*. Baltimore, Maryland: Paul Brookes Publishing Co.

Rvachew, S., Chiang, P., & Evans, N. (2007). Characteristics of speech errors produced by children with and without delayed phonological awareness skills. *Language, Speech, and Hearing Services in Schools, 38*, 60-71.

Rvachew, S., & Jamieson, D. G. (1989). Perception of voiceless fricatives by children with a functional articulation disorder. *Journal of Speech and Hearing Disorders, 54*, 193-208.

Rvachew, S., Nowak, M., & Cloutier, G. (2004). Effect of phonemic perception training on the speech production and phonological awareness skills of children with expressive phonological delay. *American Journal of Speech-Language Pathology, 13*, 250-263.

Rvachew, S., Ohberg, A., Grawburg, M., & Heyding, J. (2003). Phonological awareness and phonemic perception in 4-year-old children with delayed expressive phonology skills. *American Journal of Speech-Language Pathology, 12*, 463–471.

Rvachew, S., Rafaat, S., & Martin, M. (1999). Stimulability, speech perception skills, and the treatment of phonological disorders. *American Journal of Speech-Language Pathology, 8*(1), 33-43.

Rvachew, S., Slawinski, E. B., Williams, M., & Green, C. L. (1999). The impact of early onset otitis media on babbling and early language development. *Journal of the Acoustical Society of America, 105*, 467-475.

St-Louis, K.O. & Ruscello, D.M. (2000). Oral Speech Mechanism Screening Examination – 3rd edition (OSMSE-3). Austin, Tx: PRO-ED.

Savariaux, C., Perrier, P., & Orliaquet, J. P. (1995). Compensation strategies for the perturbation of the rounded. *The Journal of the Acoustical Society of America, 98*(5), 2428-2422.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). Clinical evaluation of language fundamentals, fourth edition (CELF-4). Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

Sherman, D., & Geith, A. (1967). Speech sound discrimination and articulation skill. *Journal of Speech and Hearing Research, 10*, 277-280.

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *J Acoust Soc Am, 125*(2), 1103-1113.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The Speech Disorders Classification System (SDCS): Extensions and lifespan reference data. *Journal of Speech, Language, and Hearing Research, 40*(4), 723-740.

Shriberg, L. D., Kwiatkowski, J., & Gruber, F. A. (1994). Developmental Phonological Disorders II: Short-Term Speech-Sound Normalization. *J Speech Hear Res, 37*(5), 1127-1150.

Smit, A. B., & Bernthal, J. E. (1983). Performance of articulation-disordered children on language and perception measures. *Journal of Speech and Hearing Research, 26*, 124-136.

Sommers, R. K., Cox, S., & West, C. (1972). Articulatory effectiveness, stimulability, and children's performances on perceptual and memory tasks. *Journal of Speech and Hearing Research, 15*, 579-589.

Stackhouse, J., & Wells, B. (1993). Psycholinguistic assessment of developmental speech disorders. *European Journal of Disorders of Communication, 28*, 331-348.

Syrdal, A. K. (1985). Aspects of a model of the auditory representations of American English vowels. *Speech Communication, 4*, 121-135.

Syrdal, A. K., & Gopal, H. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79*, 1086-1100.

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature, 423*(6942), 866-869.

Van Riper, C. (1963). *Speech correction: Principles and methods*. Upper Saddle River, NJ: Prentice-Hall.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J Acoust Soc Am, 122*(4), 2306-2319.

Wolfe, V., Presley, C., & Mesaris, J. (2003). The importance of sound identification training in phonological intervention. *American Journal of Speech-Language Pathology, 12*, 282-288.

Author Note

This work was supported by fellowships to the first author from the Center for Research on Language, Mind and Brain and the Canadian Child Health Clinician Scientist Program and a research grant to the second author from the Social Sciences and Humanities Research Council. Correspondence concerning this article should be addressed to Douglas Shiller, PhD, École d'orthophonie et d'audiologie, Faculté de médecine, Université de Montréal, C.P. 6128, Succursale Centre-Ville, Montréal, Québec, H3C 3J7. Email: douglas.shiller@umontreal.ca.

Received: June 25, 2009

Accepted: Mar 12, 2010