<table>
<thead>
<tr>
<th><strong>Author(s):</strong></th>
<th>McLeod, S.L.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title:</strong></td>
<td>The impact of seeing speech</td>
</tr>
<tr>
<td><strong>Book:</strong></td>
<td>Translational speech-language pathology and audiology: Essays in honor of Dr. Sadanand Singh</td>
</tr>
<tr>
<td><strong>ISSN:</strong></td>
<td>9781597564458</td>
</tr>
<tr>
<td><strong>Year:</strong></td>
<td>2012</td>
</tr>
<tr>
<td><strong>Pages:</strong></td>
<td>245 - 251</td>
</tr>
<tr>
<td><strong>Editor:</strong></td>
<td>Robert Goldfarb</td>
</tr>
<tr>
<td><strong>Publisher:</strong></td>
<td>Plural Publishing</td>
</tr>
<tr>
<td><strong>Place of Publication:</strong></td>
<td>San Diego, CA</td>
</tr>
<tr>
<td><strong>URLs:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>FT:</strong></td>
<td><a href="http://www.pluralpublishing.com/publication_tslpa.htm">http://www.pluralpublishing.com/publication_tslpa.htm</a></td>
</tr>
<tr>
<td><strong>PL:</strong></td>
<td><a href="http://researchoutput.csu.edu.au/?func=dbin-jump-full&amp;object_id=42893&amp;local_base=GEN01-CSU01">http://researchoutput.csu.edu.au/?func=dbin-jump-full&amp;object_id=42893&amp;local_base=GEN01-CSU01</a></td>
</tr>
</tbody>
</table>
Dr. Sadanand Singh was a visionary in the fields of speech-language pathology and audiology. While his writing and research primarily addressed phonetics, phonology, measurement, and clinical procedures, his vision for dissemination of new ideas enabled authors and researchers to extend the boundaries of these fields. This chapter embraces Dr. Singh’s interest in phonetics, instrumentation, and translational research. Dr. Singh co-authored two books with me: *Speech Sounds: A Pictorial Guide to Typical and Atypical Speech* (McLeod & Singh, 2009a) and the companion book *Seeing Speech: A Quick Guide to Speech Sounds* (McLeod & Singh, 2009b). These books were a synthesis of our work to enable speech-language pathologists (SLPs) and others to see speech. The books combined Dr. Singh’s images of speech created by acoustic (spectrographic) and cinematographic technologies with images created by electropalatography and ultrasound (see Figure 1). Static and dynamic images were presented and described for 24 consonants, 10 vowels and 5 diphthongs. The aim of this chapter is to discuss the importance of seeing speech, the technological advances that have enabled SLPs to see speech, and the impact that this can have when working with children and adults who have difficulties producing speech sounds.
The importance of seeing speech

The production of consonants and vowels to generate speech is both an acoustic and articulatory event, mediated by cognitive decision-making around the input and output of speech. In most clinical contexts, SLPs make auditory-impressionistic transcriptions of their clients’ speech using the International Phonetic Alphabet (IPA) and the extensions to the International Phonetic Alphabet (extIPA). Transcription of speech is an efficient method to document acoustic perception of speech sounds; however, as Kent (1996, p. 7) indicates, auditory-impressionistic transcription is “susceptible to a variety of sources of error and bias”. Sources of error and bias include variability due to listener characteristics, speaker characteristics, context, auditory salience, and measurement procedures. During auditory-impressionistic transcription important information can be missed, particularly when speakers have atypical or unintelligible speech:

“...highly unintelligible speech requires an exact description of the client’s speech patterns in order to plan appropriate remediation. Some kind of direct or indirect imaging technique is indicated where impressionistic transcription leaves too wide a margin of uncertainty” (Ball, Manuel & Müller, 2004, p. 161).

In addition to the transcription of speech, SLPs use their knowledge of tongue placement during assessment to categorize speech sound errors, and during intervention to describe changes in tongue placement to facilitate accurate productions. For example, the term “fronting” is often used to describe when children attempt to produce velar consonants (e.g., /k/) but produce alveolar consonants instead (e.g., [t]). For children who are fronting, intervention involves encouragement to put their tongue at the back of their mouth to achieve correct productions of /k/. In a recent survey, McLeod (2011) described 175 SLPs’ knowledge of tongue placement for speech sound production. It was found that SLPs demonstrated good knowledge of tongue/palate contact along the midline (i.e. along the
sagittal plane), but poor knowledge of contact along lateral margins of the palate (i.e. along the coronal plane). These SLPs were most accurate in their knowledge of tongue/palate contact for consonants with no contact /h, p, f/; then velar consonants /ɡ, k, ɳ/. The remaining consonants were rarely accurate, particularly those most frequently targeted in SLP intervention for children with speech sound disorders. These SLPs did not show awareness of the central groove for the fricatives /s, z, ʃ, ʒ/, lateral bracing (side contact, or “horseshoe” contact) for alveolar consonants /t, d, n, s, z/, or posterior lateral contact for most other consonants. A recommendation from this research was for SLP education to target awareness of tongue placement for consonant production (i.e. for SLPs to see speech).

/H1/Technological advances for seeing speech

Over the past 80 years, there have been numerous technological advances that have enabled speech-language pathologists (as well as others) to see speech (Ball & Code, 1997; Ball, Gracco & Stone, 2001). Early efforts to see speech relied on painting chalk and oil on the tongue and palate (e.g., Moses, 1939; Shohara & Hanson, 1941). Using these technologies tongue/palate contact for speech was identified along the coronal plane, indicating that alveolar consonants were produced with contact along the margins of the palate (more recently described as horseshoe shape contact with lateral bracing; Gibbon & Wood, 2010). Velar consonants were identified as being produced with contact across the juncture between the hard and soft palate that extended to the lateral margins of the palate. These early findings have been replicated using high-speed electropalatographic (EPG) techniques (Hardcastle & Gibbon, 1997; see McLeod & Singh, 2009a for a review of EPG images produced by speakers varying in age, language, and disorder type).

Next, acoustic technologies, using waveforms and spectrograms, were made available to see speech (e.g., Harris, Hoffman, Liberman, Delattre, & Cooper, 1958; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Dr. Singh was amongst the pioneers in the use of
speech science technology for seeing and understanding speech production and perception (Singh & Singh, 2006). Initially, acoustic technologies were only available to those with access to speech science laboratories; however, with current technological advancement, these analyses techniques now are available free of charge to be downloaded onto personal computers. One of the most widely used programs for this purpose is Praat (Boersma & Weenink, 2009). With accessibility of seeing speech via waveforms and spectrograms, researchers and SLPs can use this technology to check auditory-impressionistic transcriptions, as well as to measure speech. More recently, general imaging technologies such as ultrasound, medical resonance imaging (MRI), and electromyography (EMG) have been employed to see speech. Additional specialized instrumental techniques such as electromagnetic articulography (EMA) also have been developed (for a review, see Ball & Code, 1997).

Each of these instrumental measures enables SLPs to visualize speech by providing real time, objective, and detailed images of speech production (Ball et al., 2001; Ball & Code, 1997). In order to elucidate the hidden articulatory aspect of speech production, instrumental measures are often combined. For example, acoustic analyses and EPG were combined in order to understand differences in segment duration in the speech of people with Parkinson’s disease compared with young and aged controls (McAuliffe, Ward, & Murdoch, 2006). Moen and Simonsen (2007) combined EPG and EMA to describe differences between the Norwegian coronal stops /t, d/ since EPG provided coronal images and EMA provided two-dimensional midsagittal images of tongue movement.

Within Speech Sounds: A Pictorial Guide to Typical and Atypical Speech (McLeod & Singh, 2009a) four instrumental techniques were used to provide images of speech: spectrography, cinematography, electropalatography and ultrasound. Each of the images of /t/ in Figure 1 were created during the production of the word tat. Dr. Singh’s descriptions of the
spectrographic and cinematographic images have been included below, followed by an explanation of the electropalatograph and ultrasound images. Figure 2 provides an image of the Articulate Assistant computer screen to demonstrate the synthesis of information from acoustic, electropalatographic, and ultrasound technologies. The EPG and ultrasound images in Figure 1 were created by taking the exact midpoint of the /t/ in the word-initial position of tat.

In the spectrogram (Figure 1a), /t/ is presented in both the initial and final positions of the word tat. To use the words of Dr. Singh “… the plosive burst for the voiceless-front stop /t/ in both the initial and final positions is above 2,000 Hz… there exists a presence of aspiration noise (over 100 msec), mostly concentrated above 2,000 Hz, for the voiceless stop /t/ in the initial position. At the final position, the aspiration noise is absent and a 200-msec silence is followed by a plosive burst. The speaker has the option of releasing or not releasing the final stop.” (McLeod & Singh, 2009, p. 49).

In the cinematographic filmstrip, “the phoneme /t/ is at the initial and final positions of the word tat /tæt/. The wide lip opening starts at the first frame of the filmstrip (Figure 1b). In the third frame the tongue is seen touching the alveolar ridge. The following vowel /æ/ shows a front-low tongue position with wide and open lips, accompanied by considerable excursion of the mandible. The vowel is used in the stressed position and hence can be seen sustained from frames 7 through 13. Starting with frame 14, the tip of the tongue begins to rise and continues to do so until complete contact is made with the alveolar ridge in frame 16. The remaining eight frames show the process of complete closure to accomplish the necessary durational cue for the final-voiceless-stop consonant /t/.” (McLeod & Singh, 2009a, pp. 47-49).
The Reading EPG image in Figure 1c uses black squares to indicate tongue contact against the electrodes, and white squares to indicate no contact. The shape of the EPG image replicates the coronal image of the hard palate, with the boundaries of the image representing the edge of the teeth along the top and sides, and the lower margin indicating the juncture between the hard and soft palate. The static EPG image (Figure 1c) is reminiscent of a horseshoe, with the tongue contacting the palate along the teeth. The tongue contacted the palate across the alveolar ridge (first two rows) with lateral bracing along the sides of the teeth (first and last columns).

The static ultrasound image (Figure 1d), contains a bright white line showing the tongue surface during production of /t/. On the right of the image is the front of the tongue, although the tip is obscured because of the acoustic shadow of the jaw. The body of the tongue is raised in the oral cavity. Above the tongue is an air shadow and diagonal muscle fibres can be seen below the tongue’s surface. By combining these four images, a greater understanding of the production of /t/ is gained compared with relying on acoustic impressionistic transcription.

The impact of seeing speech in speech-language pathology practice

The ability to see speech enhances SLPs’ accuracy in diagnostic assessment and enables biofeedback during intervention for children and adults who have difficulties producing speech sounds. First, seeing speech has had an impact on SLPs’ diagnostic accuracy. For example, Gibbon (1999) reviewed EPG studies of consonants perceived as substitutions, distortions, and even correct productions. She showed that 12 (71%) of 17 school-aged children with speech sound disorders were using their tongue to cover the whole palate (undifferentiated lingual gestures). Gibbon (1999, p. 382) concluded: “Standard transcriptions do not reliably detect undifferentiated gestures, which are transcribed as speech errors … in some contexts, but are transcribed as correct productions in other contexts. Undifferentiated
gestures are interpreted as reflecting a speech motor constraint involving either delayed or deviant control of functionally independent regions of the tongue.”

Seeing speech has also had an impact on SLPs’ intervention practices, particularly in the use of instrumentation as a biofeedback device (Gibbon & Wood, 2010). Instrumentation has enabled children and adults to see their own speech, compare their speech with a target that is typical of the ambient language, and change their speech production. For example, Bacsfalvi, Bernhardt and Gick (2007) used electropalatography and ultrasound techniques for training three adolescents with severe hearing impairment to change their production of vowels. Improvements were documented using vowel formant values, EPG tongue-palate contact patterns and phonetic transcription. McAuliffe and Cornwell (2008) used EPG to enable an 11-year-old girl with a persistent lateral lisp to make consistent changes in her production of fricatives as judged by naïve listeners and documented by acoustic analysis of /s/ spectra. On a national scale, Lee et al. (2007) described how children with cleft lip and palate across Scotland benefitted from intervention using portable EPG devices to practice correct productions.

Instrumentation for seeing speech is used in research contexts, and is increasingly used in clinical contexts. Continual advancement of technologies is likely to result in less-invasive, less-expensive, and more informative technologies for seeing speech (Wrench, 2007), which in turn will further impact SLP assessment and intervention practices.

/H1/Acknowledgment

This chapter was written with support from Australian Research Council Future Fellowship (FT0990588).

/H1/References


Figure 1. Seeing speech by comparing acoustic, cinematographic, electropalatographic, and ultrasound images of /t/ (McLeod & Singh, 2009a, p.46, 48)

(a) Dynamic acoustic image (spectrogram) of /t/ in \textit{tat}

(b) Dynamic cinematographic image of /t/ in \textit{tat}

(c) Static electropalatographic image of /t/ in \textit{tat}

(d) Static ultrasound image of /t/ in \textit{tat}
Figure 2. The computer screen printout of Articulate Assistant Advanced (AAA) for combining acoustic, ultrasound, and electropalatography technologies in order to see speech in the sentence “I see a tat again” (McLeod & Singh, 2009a, p. 354)