Exploring the Use of Electropalatography and Ultrasound in Speech Habilitation

Explorer l'électropalatographie et l'échographie pour l'éducation de la parole

Barbara Bernhardt, Penelope Bacsfalvi, Bryan Gick, Bosko Radanov and Rhea Williams

Abstract

Electropalatography (EPG) and ultrasound have been recently explored as articulatory visual feedback tools in speech habilitation at the University of British Columbia's Interdisciplinary Speech Research Laboratory (UBC, ISRL). Although research studies imply that such tools are effective in speech habilitation, most studies have utilized trained listeners. To determine the impact of speech habilitation on everyday communication, it is also important to include everyday, untrained listeners in the treatment evaluation process. Two everyday listener studies were conducted, using data from two of the exploratory UBC treatment studies. The listeners observed improvement post-treatment for some but not all speakers or speech targets. More research is needed to determine the relative effectiveness of EPG and ultrasound in speech habilitation in terms of speaker variables and treatment targets, and in comparison with each other and different treatment methods. The current paper has two purposes: (1) to provide an overview of EPG and ultrasound in speech habilitation, and (2) to present the two preliminary listener studies, suggesting directions for future research and clinical application.

Abrégé:

Le laboratoire de recherche interdisciplinaire sur la parole de l'University of British Columbia (UBC) a examiné la possibilité d'utiliser l'électropalatographie et l'échographie comme outils de rétroaction visuelle de l'articulation. Bien que des études laissent entendre l'utilité de tels outils pour l'éducation de la parole, la plupart sont fondées sur des auditeurs formés. Pour évaluer l'efficacité de la réadaptation de la parole dans la communication quotidienne, il est aussi important d'inclure des auditeurs ordinaires n'ayant pas été formés au processus d'évaluation du traitement. Le laboratoire a mené, à UBC, deux études avec des auditeurs ordinaires à partir des données de deux études exploratoires portant sur le traitement. Les participants ont noté des améliorations à la suite du traitement, mais pas chez tous les orateurs ni pour toutes les cibles. Il faut poursuivre la recherche afin de déterminer l'efficacité relative de l'électropalatographie et de l'échographie pour la réadaptation de la parole en fonction des variables des orateurs mêmes et entre les méthodes de traitement. Le présent article vise deux objectifs: (1) effectuer un survol de l'utilité de l'électropalatographie et de l'échographie dans la réadaptation de la parole, et (2) présenter les résultats de deux études préliminaires sur des auditeurs afin de proposer des orientations pour la recherche et l'application clinique.

Key Words: electropalatography, ultrasound, everyday listener, visual feedback

Barbara Bernhardt, *Ph.D.*
School of Audiology and Speech Sciences
University of British Columbia
Vancouver, BC Canada

Penelope Bacsfalvi, *M.Sc.*
School of Audiology and Speech Sciences
University of British Columbia
Vancouver, BC Canada

Bryan Gick, *Ph.D.*
Department of Linguistics
University of British Columbia
Vancouver, BC Canada

Bosko Radanov, *B.A.*
School of Audiology and Speech Sciences
University of British Columbia
Vancouver, BC Canada

Rhea Williams, *M.Sc.*
Speech-Language Pathologist
Barrie, ON Canada

In the past several decades, a number of small *n* studies have demonstrated the utility of visual feedback technology in speech habilitation. Studies have included participants with a variety of etiologies, for example, hearing impairment (e.g. Bernhardt, Fuller, Loyst & Williams, 2000; Bridges & Huckabee, 1970; Dagenais, 1992; Fletcher, Dagenais, & Critz-Crosby, 1991; Volin, 1991), cleft palate (e.g., Gibbon, Crampin, Hardcastle, Nairn, Razzell, Harvey, & Reynolds, 1998; Michi, Yamashita, Imai, & Yoshida, 1993), phonological impairment (e.g., Adler-Bock, 2004; Gibbon, Hardcastle, Dent, & Nixon, 1996), or motor speech impairment (e.g., Gibbon & Wood, 2003). Technologies providing acoustic displays are perhaps most common (e.g., Maki, 1983; Shuster, Ruscello, & Toth, 1995; Volin, 1991). However, there is a growing interest and body of research on visual articulatory feedback in treatment, for example, electropalatography (EPG, or palatometry), which shows dynamic tongue-palate contact patterns (e.g., Bernhardt et al., 2000; Dagenais, 1992; Fletcher et al., 1991; Gibbon et al., 1998; Hardcastle, Jones, Knight, Trudgeon, & Calder, 1989). Two-dimensional ultrasound, which can display dynamic images of tongue shapes and movements, has also become more accessible in the past 5 years for speech habilitation (Adler-Bock, 2004; Bernhardt, Gick, Bacsfalvi, & Ashdown, 2003; Gick, 2002). At the University of British Columbia, exploratory treatment studies have been conducted using EPG and/or ultrasound. One purpose of the current paper is to provide an overview of these technologies in speech habilitation as a foundation for future clinical and research applications. Although the tools are currently university-based, there may be potential for future clinical use. Queen Margaret University College in Edinburgh has been situating EPG in clinical sites throughout Britain, with links to the university research team (Cleftnet UK: Gibbon et al., 1998). One step in proceeding towards such a program in Canada is dissemination of information about the technologies and their clinical use to S-LPs and other researchers. For clinical purposes, the question is whether technologies are worth the investment. The research literature suggests that visual feedback technologies aid speech habilitation. However, most studies to date have been conducted by trained observers. Because an ultimate goal of speech therapy is to enhance communication in a client's everyday life, outcomes research also needs to include the perspectives of everyday listeners (Frattali, 1998; Klasner & Yorkston, 2000; World Health Organization, 2001). Two of the exploratory UBC treatment studies had data that could be used to collect everyday listener observations. The second purpose of the current paper is to present those preliminary everyday listener observations, not as measures of effectiveness of the treatment, but as a foundation for future research and clinical studies. By discussing the two small studies in one paper, a broader perspective can be gained on everyday listener research.

The paper begins with an introduction to EPG and ultrasound and a brief discussion of everyday listener research methods. The two listener studies are then presented in turn. Background information is provided on the treatment studies themselves within the context of each listener study, although space precludes a detailed discussion of them (see Bernhardt et al., 2000 and Bernhardt et al., 2003 for more details). The treatment studies were case studies conducted with the purpose of learning about the use of EPG and ultrasound in speech habilitation. Thus, they were conducted without strict experimental single subject or group designs (e.g., the use of control groups). The projects were developmental in nature, and thus, S-LPs and researchers shared information with each other throughout about participants and procedures. The final section of the paper discusses future research and clinical implications.

Visual Feedback Technology

Dynamic Electropalatography

Different types of dynamic EPG systems have been available over the past three decades. Older systems such as the now unavailable Kay Palatometer ran on DOS (Kay Elemetrics, New Jersey), with more recent ones running on Windows, for example, the WIN-EPG (www.articulateinstruments.co.uk) or the Logometrix system (www.logometrix.org). The Kay palatometer and the WIN-EPG (2000) have been used in the UBC research program, and thus the discussion focuses on these instruments.

The above-mentioned systems operate in similar ways. Speakers wear a custom-fit artificial palate (figure 1).

Figure 1

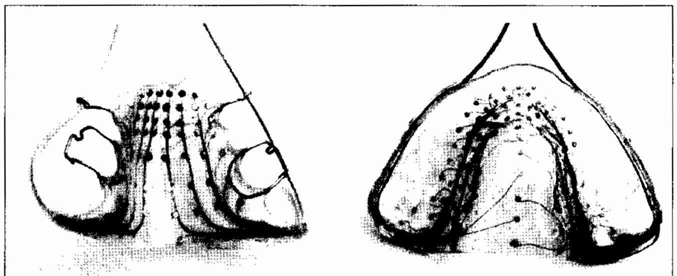


Figure 1. The WIN-EPG and Kay artificial palates: The top of the palates represents the front part of the mouth, and the bottom, the velar area. The WIN-EPG has 62 electrodes bunched densely at the front of the palate, but with contacts across the palate to the velar area. The Kay pseudopalate has 92 electrodes, bunched densely around the edge of the palate up to the teeth, but with contacts back to the velar area.

Although highly anomalous oral structures may preclude the wearing of artificial palates, all 22 speakers in the UBC projects have been able to wear them. The palate contains electrodes that are sensitive to tongue-palate contacts and send contact-induced electrical impulses through fine bundled wires to a computer. Tongue-palate contact patterns are displayed on-line. These displays and the accompanying acoustic signals can be stored for analysis, or as templates for practice. The Kay and WIN-EPG differ in how the palates attach, the number and type of electrode displays and the types of analyses provided. The acrylic Kay pseudo-palate has 92 electrodes; it fits over and is held on by the upper teeth (figure 1). The acrylic WIN-EPG has 62 electrodes; dental wires hold the pseudo-palate onto the upper teeth. Both types of palates allow distinction of non-low vowels and alveolar to velar consonants from one another; the Kay palate, in addition, provides displays of dental consonants. The EPG systems typically provide acoustic displays (e.g., a waveform displaying intensity), with the WIN-EPG also providing off-line spectrographic displays and detailed quantitative analyses of contact patterns.

The following section describes typical maximal target contact patterns and some aberrant patterns for English lingual consonants and vowels, as observed by the authors during the treatment studies (see figures 2a-2h, 5a-d and 6a-d). EPG images for the current paper are taken from the currently used machine, WIN-EPG, which has much easier image exporting capabilities than the previous machines. Although exact tongue-palate contact patterns for speech targets vary within and between speakers, the contact patterns are similar in configuration and region. Alveolar targets /t/ (figure 2a), /d/ and /n/ show a horseshoe contact pattern, with the tongue touching the alveolar ridge and the sides of the upper dental arch to the back of the molar region. The alveolar sibilants /s/ (figure 2b) and /z/ show a similar contact pattern, but have more contact on the sides of the palate, creating a groove primarily in the alveolar area. This groove varies somewhat in size and location across speakers (McLeod, Roberts & Sita, 2003). The contact pattern for /l/ depends on context (figure 2d). Prevocalic /l/ generally has alveolar contact similar to /t/,/d/ and /n/, full or near-full lateral contact on one side of the upper dental arch, and reduced posterior lateral contact on the other side of the dental arch. Postvocalic "dark" /l/ tends to show posterior velar contact at the beginning of the articulation, followed rapidly by the contact pattern for prevocalic /l/. The palatoalveolars /ʃ/ (figure 2c) and /ʒ/ show broad, post-alveolar contact along the sides of the upper dental arch, and a wider groove (i.e., less contact) than the alveolar sibilants. The affricates /tʃ/ and /dʒ/ vary across speakers. Some speakers show a stop-sibilant contact pattern in the palatoalveolar area (Hardcastle, Gibbon & Scobbie, 1995), while others show movement backwards from the /t/ or /d/ contact area to a post-alveolar /ʃ/ or /ʒ/ contact area. English /r/ tends to have lateral contact along the back molars and a wide channel with no contact

Figure 2

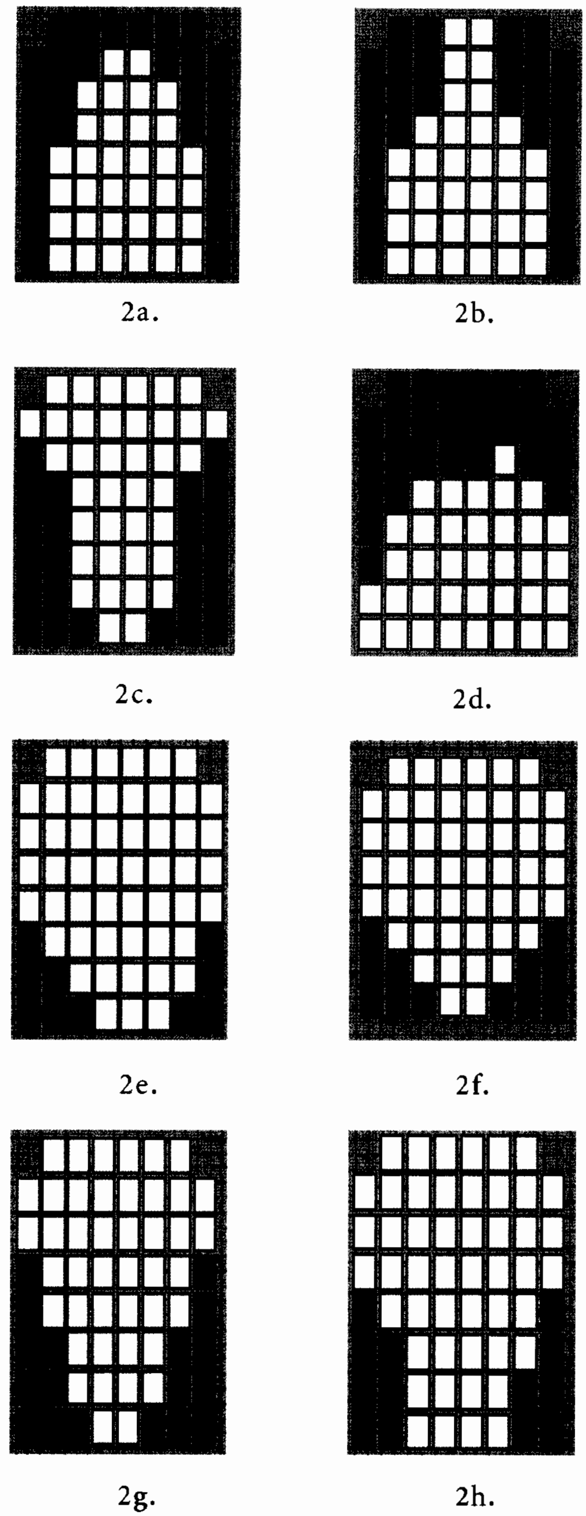


Figure 2a-2h. Electropalatograms for North American English (a) /t/, (b) /s/, (c) /ʃ/, (d) prevocalic /l/, (e) /k/, (f) prevocalic /r/, (g) /i/, and (h) /ɪ/. The top of the figure represents the alveolar area, and the bottom, the velar area. The black squares indicate tongue-palate contact.

in the middle of the tongue (figure 2f). The velar consonants (figure 2e) tend to have continuous contact along the back of the pseudo-palate region in the context of back vowels, although for some speakers, this contact is not visible if their pseudo-palate does not extend far enough back. In the context of front vowels, the velars show continuous contact from the velar through the palatal regions. The front vowels and /j/ show bilateral contact about halfway along the palate towards the front, and a wide mid-channel with no contact; the back vowels and /w/ show minimal contact in the velar region and a wide mid-channel with no contact (figure 2g, 2h). Tense vowels have the same type of contact pattern as their lax vowel cognates, but are produced further forward along the dental arch. Low vowels are generally not visible because they have no tongue-palate contact in most speakers; for some speakers, mid vowels have no visible contact.

During speech therapy, the target tongue-palate contact pattern is demonstrated by a typical speaker with an artificial palate, and the client is encouraged to emulate that target. If the client produces an acceptable variant of the target phone with a different tongue-palate contact configuration from that of the model, the client's production becomes his or her own template.

For further information on EPG, consult http://sls.qmuc.ac.uk/epg/epg0_big.html.

Dynamic Two-Dimensional Ultrasound

The following description provides an overview of the functioning of two-dimensional ultrasonography for speech displays. To display speech or other lingual movements with ultrasound, a transducer is held by the speaker or attached to a fixed arm or stand so that it contacts the undersurface of the speaker's chin. The transducer is coated with water-soluble ultrasound gel to enhance the signal. Sound waves are transmitted by the transducer up through the oral tissue. Echo patterns from sound waves returning from the tongue surface are converted to moving images that are visible on an ultrasound screen (see figures 3, 4 and 7). (See also Stone, 2005.) There are a number of different dynamic two-dimensional ultrasound machines, with large differences in price, depending on the size and complexity of the machine and transducer. Three different machines have been used in the UBC treatment studies. For Study #2 in the current paper, the Aloka SSD-900 portable ultrasound was used, along with a 3.5MHz convex intercostal transducer probe. The range and gain were adjusted to give the clearest image for the tongue surface across assessments, for example, a range of 11 centimeters, gain of 60. The simultaneous audio signal was recorded onto VHS tape at 30 frames per second from the ultrasound machine (JVC Super VHS ET Professional Series, SR-VS20), and also onto a digital video source using a Pro-Sound YU34 unidirectional microphone amplified through the built-in pre-amplifier in a Tascam cassette deck. In subsequent studies, two other machines have been utilized: (a) a portable Sonosite 180 Plus with a Sonosite C15/4-2 MHz MCX transducer (for treatment only), and (b) a stationary Aloka Pro-Sound SSD-5000 with a UST-9118 endo-vaginal 180-degree convex array transducer (see Bernhardt, Gick, Adler-Bock & Bacsfalvi, 2005).

Figure 3

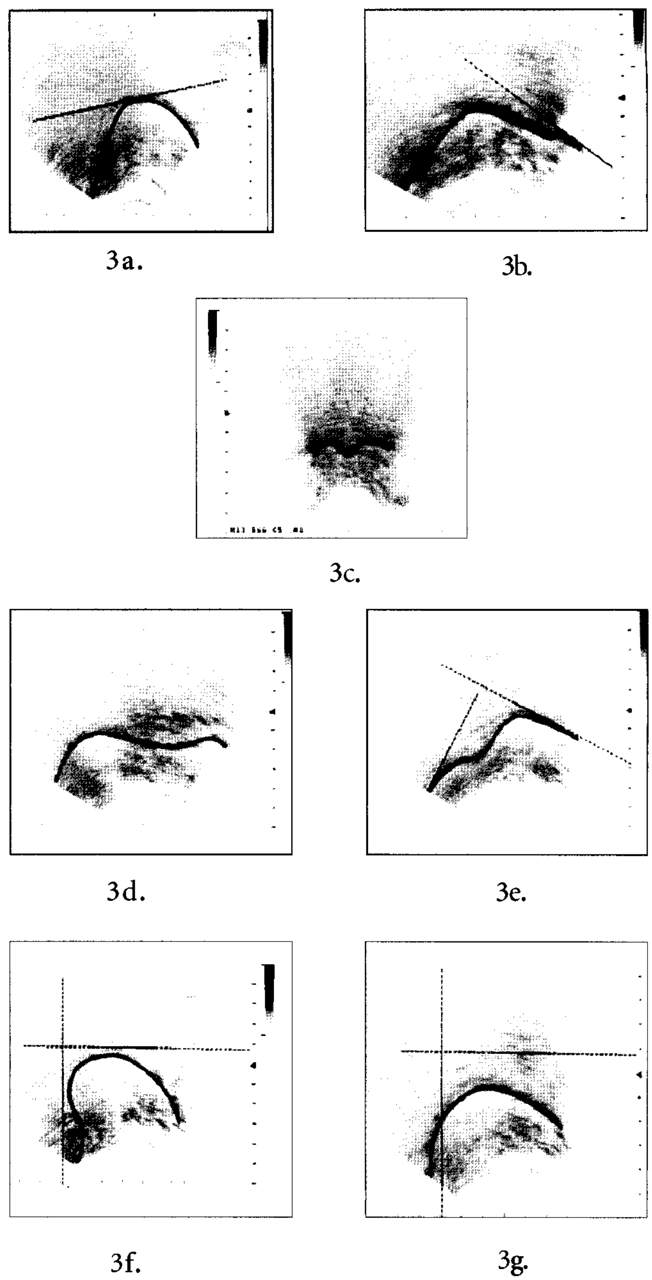
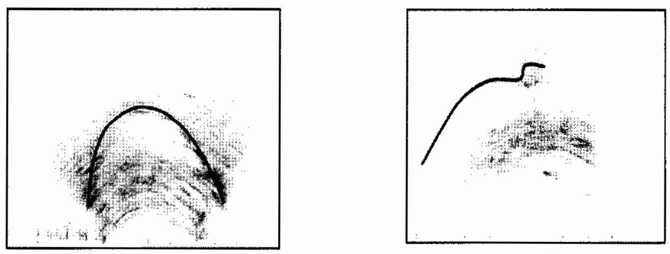


Figure 3a-3g. Ultrasound images of North-American English: (a) /k/, (b) /t/, (c) /s/ (coronal view), (d) /l/, (e) /r/ (f), /u/, (g) /ʊ/. The tongue tip is on the right in all sagittal images. The straight line for /k/ approximates the velar area. The straight line for /t/ approximates the alveolar area. 3c shows a mid-line groove for /s/. Note the complex shapes of /l/ and /r/ (3d, 3e). The /u/ has an advanced tongue root and higher tongue body than /ʊ/ (3f, 3g).

Either mid-sagittal (figure 3a,3b,3d,3e,3f,3g) or coronal-oblique (figure 3c) views of the tongue shape and movement patterns can be displayed on a screen. The sagittal view displays a side view of the tongue, showing tongue height, backness, and slope; this view has been used most in the UBC studies. The tip or root may be obliterated in the display, due to the limited field of view or by the jaw and hyoid shadows. However, the general shape, position, height and slope provide useful feedback to the speaker. Slope has been found to be especially relevant for /l/, /r/ and the vowels in the ISRL studies. The coronal view shows a cross-section of the tongue, and thus, flatness, mid-line grooving or lateral elevation or depression of the tongue. This view has been useful for showing mid-line tongue grooving for sibilants and /r/, plus relative vowel height. Reference points or lines or palatal contour sketches derived from images of swallows are sometimes added to the display. This is done either by attaching overhead transparencies with drawings of the palatal arch or target tongue positions or shapes to the screen, or by using reference lines generated by the ultrasound machine itself. For example, if a certain tongue height is optimal for a particular target, a reference line will be placed on the screen that the speaker is to 'hit' with his or her tongue body or tip. (See the lines in figures 3a, 3b and 3e, for example.) As with EPG treatment, a typical speaker provides examples for the client to emulate. Images or movies can be stored and used for future reference.

Typical and aberrant ultrasound images are shown in figures 3, 4 and 6 (see also Bernhardt et al., 2005), and are described below. The tongue tip movement and height for the velars contrast visibly with tongue body movement and height for the alveolars, as shown in figures 3a and 3b respectively. For sibilants, the varying width of the groove for the alveolar and post-alveolar sibilants is visible in the coronal view as in figure 3c. The sagittal view shows the relative front-back position of the apical end of the tongue and helps distinguish alveolar from post-alveolar fricatives. For affricates, the display shows the relative backness of the tongue and any change in movement from a more anterior position for the stop portion of the affricate to a more posterior position for the fricative portion. The English /l/ and /r/ are complex articulations with multiple components that differ across word position and speakers (figure 3d and 3e; also see figure 4 which shows a sample aberrant /r/ pre-treatment and accurate /r/ post-treatment from Adler-Bock, 2004). For both liquids, the sagittal view typically shows a two-point displacement of the tongue in the tip/blade and root regions (Stone & Lundberg, 1996). For /l/, the relative timing of the anterior and posterior constrictions can also be seen in the sagittal view (figure 3d; Sproat & Fujimura, 1993; Gick, 2003). The coronal view shows the lateral dip of one or both sides of the tongue for /l/, a dip which is usually towards the posterior portion of the tongue body. The English /r/ can be produced with a more retroflexed or bunched position, and is generally articulated with three separate constrictions along the vocal tract (Delattre & Freeman, 1968): labial, central and pharyngeal. The shape of the tongue in the region of each of the lingual constrictions is visible on ultrasound (figure 3e). A posterior and relatively wide mid-line depression is another important component of /r/ and is visible in the coronal view of /r/. With respect to vowels, the sagittal view provides a view of the whole tongue as it advances and retracts, and moves through various heights. The sagittal view thus displays the tongue root as it advances and retracts for tense and lax vowels respectively, and in addition, shows the higher tongue body position for tense vowels (compare the height and backness of /u/ and /ʊ/ in figure 3f and 3g, and the same for /i/ and /ɪ/ in figure 7). The coronal view also shows relative height of the sides of the tongue; the sides are higher for the high and tense vowels. Additional information on ultrasound is available in the Volume 19, 2005 issue of *Clinical Linguistics and Phonetics* and the following websites: www.linguistics.ubc.ca/isrl/UltraSoundResearch/; http://speech.umaryland.edu/research.html; www.slp.utoronto.ca/People/Labs/TimLab; www.sls.qmuc.ac.uk/RESEARCH/Ultrasound.

Figure 4



4a. 4b.

Figure 4a, 4b. Ultrasound images of a pre-treatment vocalic substitution for /r/ (4a) followed by a post-treatment on-target /r/ production (4b: Victor, Adler-Bock, 2004).

Everyday Listener Observations

An ultimate goal of speech therapy is to enhance communication in a person's everyday life. The observations of everyday listeners are thus important in the treatment evaluation process. A variety of methods can be used to obtain such observations, from studies of speech intelligibility or comprehensibility to more qualitative approaches such as interviews, questionnaires or focus groups. Speech intelligibility measures were used in the two following studies; consequently, the following discussion focuses on those methods.

Speech intelligibility is generally evaluated with some kind of identification task or rating scale (Kent, Miolo & Bloedel, 1994). For identification tasks, words may be presented in isolation or in connected speech in a variety of listening conditions. Listeners may be asked to transcribe orthographically what they hear, or to select responses from closed sets in computerized or non-computerized protocols. Kent et al. (1994) suggest that identification tasks can provide information about specific words and phonemes, but not about the speaker's general conversational competence. General scalar ratings of connected speech samples, in contrast, can provide holistic appraisals of speech, because the listener can take non-segmental factors into account, such as intonation, rate, and rhythm. The general rating scales, however, give minimal information on specific words and segments. Rating scales may also be more subject to listener and context biases than identification tasks (Schiavetti, 1992; Kent et al., 1994). Some rating scales may provide more information on specific speech targets (Black, 1999; Ertmer & Maki, 2000). Ertmer and Maki note that progress in treatment often has an intermediate phase, in which "closer—though still distorted approximations" may precede fully acceptable variants of the target (2000, p. 1514). They constructed a 3-point rating scale for evaluating treatment of specific targets: "0" (omission or substitution of the target), "1" ("improved but not fully acceptable"), or "2" ("a fully acceptable" variant of the target) (p. 1514). This method also has inherent biases. First, the speech target is known, potentially influencing the listener responses. Furthermore, listeners may vary in their definition of acceptability, depending on their background and their understanding of the task. Nevertheless, the task does allow specific speech targets to be rated without phonetic transcription, making it usable by everyday listeners.

One of each type of task was selected for the current listener studies in order to assess the utility of each for evaluating data from treatment studies. The first study used a single word identification task. The second study adopted a 3-point rating scale similar to the Ertmer and Maki (2000) scale. Based on formal (Bernhardt et al., 2003) and informal trained listener observations, it was predicted that everyday listeners would identify more post-treatment words in study #1, and rate post-treatment speech samples more highly in study #2. It was also predicted that their responses would differ according to the various speakers. However, it could not be predicted to what degree pre- and post-treatment listener observations might differ, or how listeners might react to individual speakers or segment (phoneme) types in the two studies. These preliminary listener responses would serve as a foundation for future research questions and methods. The two studies are discussed in turn below.

Study #1: Speech Habilitation Using EPG with a Heterogeneous Group of Speakers

Treatment Study Background for Listener Study #1

The first treatment study included a heterogeneous sample of 7 speakers in terms of etiology, age and speech target types (see table 1). Inclusion of a variety of speaker types in the first exploratory study allowed the research team to gain preliminary insight into the overall scope of the technology for future studies. Three speakers had hearing impairments, 4 had cleft palates, and 2 had motor impairments, thus reflecting the variety of impairments reported in the EPG literature. Four speakers were 18 or over, and three were 8-9 years of age. Two adults had mild speech impairments (pseudonyms Stan and Delia), one had a moderate speech impairment (Devon), and one a severe speech impairment (Samantha). Among the children, 2 had severe speech impairments (Dora, Sandy), and one, a moderate speech impairment (Dana). All speakers had received a minimum of 3 years of prior speech therapy.

Phonological analysis (Bernhardt & Stemberger, 2000) and the palatograms served as a basis for setting individualized targets for treatment. Speech targets included: (1) alveolars /t/, /d/, /s/, /l/, with all speakers having at least one alveolar target, (2) palatoalveolars /ʃ/ and/or /tʃ/ and /dʒ/ for Stan, Delia, Dora, Devon, (3) velars /k/ and/or /g/ for Delia, Samantha, Sandy, Devon, (4) /θ/ for Sandy and Delia, and (5) /r/ for Dora and Samantha. One of two S-LPs associated with the project conducted a traditional articulation therapy baseline (Bernthal & Bankson, 2004) over 6-8 sessions, focusing on one or more of the speakers' targets. The S-LPs modeled targets for imitation, provided visual, auditory and tactile cueing, and gave oral, written or sign language feedback as indicated. For the children, games and books were used to engage their interest. The children's and Devon's family members attended the sessions. Narrow phonetic transcription by the investigators at the end of the baseline period showed no change for identified targets.

A 20-session palatometry program was then conducted over 14-16 weeks at the university by one of the same two S-LPs in consultation with the first author. The case study protocol for each speaker consisted of two 4-week treatment blocks with two treatment sessions per week, a 1-to-3-week treatment break, and a 4-session weekly maintenance phase. More than one target was included for each client in each block in a semi-cyclic approach to treatment. The training time for specific targets was adjusted to speaker needs, with targets being revisited in the second treatment block as required. The palatometer was considered an adjunct to the articulation therapy program, which was conducted with the same general therapy techniques used in the baseline. Participants practised the given targets as isolated segments, and then in syllables, words, sentences and conversation, both with and without the palatometer. As the client progressed at each level of complexity (e.g., segment, syllable, etc.), targets were practiced more often without the artificial palates within sessions. From the beginning, home practice was encouraged without the artificial palates for those targets that showed some success in the sessions. For example, a speaker might be asked to practice oral movements, segments, syllables, words or phrases, with a family member providing feedback.

Table 1

Study #1 speaker characteristics and EPG treatment targets

Speaker's pseudonym	Age	History	Major speech patterns	Severity and EPG targets
Stan	19	Cleft lip and palate	Alveolars and palatoalveolars: mid-palatal substitutions. Sibilants lateralized. Mild nasal emission	Mild /t, s, st, ʃ, dʒ/
Delia	29	Cleft palate	/t/, /d/, /k/, /tʃ/ : glottalized Sibilants ungrooved, often pharyngealized.	Mild /θ, t, d, s, z, k, g/
Devon	40	Cerebral haemhorrage.	Imprecise articulation, voicing mismatches, hypernasility, suprasegmental aberrations	Moderate /t, d, s, ʃ, k, g/
Samantha	18	Rubella: Profound hearing loss, mild oral-motor left-sided weakness, palatal lift. ESL: Oral (some sign).	Many deletions, substitutions. Nasal emission without palatal lift. Weak articulation	Severe /s, ʃ, k, r/
Dana	8	Cleft lip and palate. Fistula. Malocclusion. Pharyngoplasty, speech bulb reduction.	Alveolars > mid-palatal or interdentals. /ʃ, /r/, prevocalic /l/: velarized	Mild /t, d, s, l/
Sandy	9	Klippel-Feil Syndrome. Cleft lip and palate Fistula. Malocclusion. Mild-moderate hearing loss (aided).	Velar stops, fricatives, affricates > glottal stops, pharyngeal fricatives, nasal fricatives Alveolar stops dentalized. Labials accurate. Prevocalic /r/ > glide.	Severe /θ, s, k, g/
Dora	8	Cytomegalovirus: Cochlear implant age 3;11. Total communication since preschool.	Liquids, some final consonants, cluster elements: deleted. /s/, /z/ : weak or deleted. Velars > alveolars. /l/ > [m]; /n/ > [f].	Severe /t, d, s, ʃ, tʃ, g, r/

Everyday Listener Study #1

Method for everyday listener study #1.

Williams (fifth author) conducted the everyday listener study described below as part of her master's thesis project (Williams, 1998). She organized the stimuli, ran the experiment and conducted initial analyses. The first author conducted further analyses for the present paper.

Eight male and eight female adult listeners participated. These listeners had a mean age of 25, Grade 12 or higher, spoke English as their first language, and had had no prior experience with disordered speech. Their hearing was assessed as normal with a pure tone screening level of 20dbHL from 500-6000 Hz, speech reception thresholds at 20dbHL or better, and speech discrimination scores of 88% or better.

The listener task involved open-set word identification. Word stimuli for the task came from two audio-recordings of each speaker pre- and post-treatment, made with a Marantz audiotape recorder and PZM 33-1090B microphone in the therapy room. Because there were no observable changes in the baseline period, the listener data included the pre- and post-treatment assessment samples only, in order to reduce listener time requirements. The stimuli were taken from a list of 164 single words not used in therapy that included multiple exemplars of English consonants and clusters across word positions (Bernhardt, 1990). A unique list of stimuli words was selected semi-randomly for each speaker. Each speaker's list contained the same pre- and post-treatment words, and seven to ten consonant treatment targets. Different words were chosen for each speaker, because listeners could potentially have learned words from the more intelligible speakers during the task, enabling them to identify those words when spoken by less intelligible speakers. Ten was the maximum number of words per speaker that could be selected, in order to include all of a speaker's treatment targets, while avoiding repetition of words across speakers.

Because of the individual case study design of the treatment study, and the heterogeneity of the speaker group, listeners were asked to make within-speaker judgments. This may have resulted in a practice effect, that is, some within-speaker familiarization for the listeners. The practice effect was diminished by randomizing stimuli within the pre- and post-treatment counterbalancing speaker order across listeners, and presenting all pre- and post-treatment samples in the first of two listening sessions. The latter procedure was considered potentially less biasing than having randomized pre- and post-treatment words from the same speaker in the same small word sets, where clearer, post-treatment pronunciations could potentially give the listener cues to pre-treatment words. A 2-3 day interval between testing sessions further reduced potential practice effects.

The audiorecorded stimuli were digitized using the Computerized Speech Research Environment 45 (CSRE45) software (1995) and the Tucker Davis Technologies (TDT) hardware (1994) at a sampling rate of 20 kHz. Sound files were edited and analyzed in the Ecoscon program in CSRE45. The sound files were attenuated or amplified during pre-processing so that pre- and post-treatment stimuli pairs could be presented at similar levels of intensity (Williams, 1998). Fourteen blocks were designed within Ecosgen, a stimuli presentation protocol that is part of CSRE45. Seven blocks contained the pre-treatment words (one block per speaker), and seven the corresponding post-treatment words. Stimuli were presented to the listeners using Ecosgen via the TDT. Participants listened through Madsen TDH 39P 10W headphones in a double-walled Industrial Acoustical Company (IAC) sound booth. During the task, listeners faced the computer screen, which showed one large square labeled 'NEXT'. Listeners were told that there were seven blocks in each session, that all stimuli in a given set came from one speaker, and that they could decide when to proceed to the next word. They were asked to write down the words that they heard. Each listening session lasted 60-80 minutes, with breaks on request.

Results and discussion of listener study #1.

Analyses were performed within-speaker only, because of the small size and heterogeneity of the sample. (See table 2.) The prediction had been that listeners would identify significantly more words and target segments in the post-treatment word sets across speakers. Listener responses varied by speaker. For pre-treatment stimuli, average words correctly identified across listeners showed a bimodal split. For speakers with a mild-moderate impairment (Stan, Delia, Devon and Dana), listeners identified an average of 6 to 7 of 10 words per speaker. For speaker with a severe impairment (Samantha, Dora and Sandy), listeners identified 0 to 2 of 10 words per speaker. The average number of correctly identified words was higher for all speakers in the post-treatment samples but to different degrees. The non-parametric Wilcoxon's Signed Ranks test was used to compare the pre-post listener

Table 2

Average speaker words and treatment targets identified across the 16 listeners in pre-and post-treatment samples.

Speaker	Avg. pre-tx words identified (/10)	Avg. post-tx words identified	Avg. treatment target segments identified pre-tx	Avg. treatment target segments identified post-tx
Stan	6.63 (0.96)	6.94 (1.61)	9.23/10 (O.66)[a]	9.54/10 (0.74)*
Delia	7.44 (0.96)	7.88 (0.50)	6.25/8 (0.45)	6.0/8 (0.85)
Devon	6.63 (1.30)	7.44 (1.15)	6.38/8 (1.02)	6.93/8 (1.16)
Samantha	0.38 (0.50)	0.75 (0.68)	0.81/9 (0.66)	2.07/9 (1.03)**
Dana	6.50 (1.67)	9.44 (0.51)**	8.25/10 (1.81)	9.87/10 (0.35)**
Sandy	2.44 (1.26)	3.63 (1.26)*	1.34/8 (1.13)	3.27/8 (0.96)
Dora	1.06 (0.57)	5.44 (1.15)**	0.63/10 (1.09)	6.07/10 (1.16)**

Note: tx = treatment. Parentheses = standard deviation. Each number represents the average number of words or segments identified across the 16 listeners out of the set of 8-10 potential words or segments.

[a]Treatment targets varied in number from 7 to 10 per speaker within word sets.

*Wilcoxon's Signed Ranks: $p < .05$

**Wilcoxon's Signed Ranks: $p < .007$ (from .000 to .006) observation sets, because of the small size and heterogeneity of the sample. The adolescents and adults showed a small, non-significant increase, although for Devon, the increase approached significance (p = .053). The children showed a significant increase: to 9.44/10 for Dana (p = .001), 5.44/10 for Dora (p < .001), and 3.63/10 for Sandy (p = .013). Standard deviation for words identified ranged from 0.57 to 1.67 words across listeners, suggesting that listeners were in close agreement. In terms of consonant treatment target identification, there was a similar split by speaker across listeners. For pre-treatment stimuli, the average numbers of target consonants correctly identified across listeners were as follows: (1) for Stan, Delia, Devon and Dana: over 75% (from 6.25/8 to 9.23/10), and (2) for Samantha, Dora and Sandy: 0.63%-15% (from 0.63/10 to 1.34/8). Post treatment, more segments were correctly identified for all speakers but Delia, for whom there was a slight non-significant decrease in consonant identification. The increase was significant according to a Wilcoxon's Signed Ranks test (p < .05) for all but Devon, who showed a near-significant increase (p = .06).

The small sample, case study design and preliminary nature of the listener study preclude in-depth statistical analyses or interpretation of the data for the various speakers. The slight and non-significant increases in word identification for Stan, Devon, Samantha and Delia may have reflected a listener practice effect for those speakers, rather than an actual improvement. Listeners did identify significantly more target consonants in the two adolescents' words (Stan and Samantha), which may however be a reflection of positive treatment effects. The increase in number of words and segments identified for the children suggests a change above and beyond a practice effect, which may or may not be attributable to the program. The differential response of the listeners to various speakers suggests that future studies will need to include larger numbers of speakers of different ages, etiologies, severity and phonological profiles.

The listeners' differential responses also show that everyday listeners can contribute useful data about individuals in an evaluation process. In terms of the listening tasks, word and segment identification can give specific information that may be useful in determining treatment effects. The issue of practice effects is a challenge in listening tasks. Randomization of stimuli may help diminish such effects, but even randomized stimuli may influence each other in small word sets, suggesting that larger word sets are needed in future studies. The results from this listener study suggested further exploration of visual feedback technology was warranted.

Study 2: Speech Habilitation using EPG and Ultrasound with Adolescents with Hearing Impairment

An Interdisciplinary Speech Research Laboratory (ISRL) was funded at UBC in 2001, making new equipment available for treatment studies, including two-dimensional ultrasound and WIN-EPG (2000). A follow-up exploratory project was initiated incorporating both EPG and ultrasound, not to compare the two technologies experimentally, but to learn more about each new technology as a basis for future studies.

Treatment Study Background for Study #2

The first treatment study had shown a difference between child and adult participants, with the adolescents (Stan and Samantha) showing slightly greater treatment effects than the other adults. In order to gain more insight into the relevance of age, four adolescents aged 16-18 were recruited for the second treatment study.

The students for the second study were a homogeneous group in terms of age and etiology (hearing impairment). The three males had aided hearing levels in the moderate range (Palmer) or moderate-to-severe range (Purdy, Peran). The female participant, Pamela, had a fluctuating and progressive sensorineural hearing loss due to Large Vestibular Aqueduct Syndrome. Her aided thresholds up to 2000 Hz sloped downwards from normal to the mild loss range in the better ear. In the other ear, her aided thresholds were in the moderate to severe loss range. Across speakers, audio-recorded baseline speech samples showed mild to moderate suprasegmental aberrations in terms of voice quality, intonation, nasalization, loudness and/or pitch control. Sibilants and liquids were the least well-established consonant categories. Vowels tended to be centralized and/or lowered, and the tense-lax distinction for vowels was only weakly established. Pamela and Purdy were intelligible most of the time in conversation; Peran and Palmer were intelligible some of the time in conversation. The adolescents attended an oral high school program for deaf and hard of hearing with partial mainstreaming and speech-language support. The adolescents had received at least 12 to 15 years of prior speech habilitation. The second author was their current speech-language pathologist. Prior to the visual feedback treatment study, this S-LP conducted a 5-week traditional treatment baseline, targeting /l/ and sibilants. Speakers showed slight gains on consonants that they could already pronounce on occasion pre-treatment. (See Bernhardt et al., 2003, for more details.)

The speech habilitation study used both WIN-EPG and ultrasound. There were 14 weekly individual treatment sessions at the ISRL, with follow-up sessions at the school without the use of technological feedback. All speakers had the same treatment targets. Consonant treatment targets included the voiceless coronal fricatives /s/ and /ʃ/ and the approximants /l/ and /r/. For vowels, the tense-lax distinction was targeted in the pairs /i/-/ɪ/ and /u/-/ʊ/. Additional data were collected as controls: /k/ and alveolar stops, /tʃ/, and all other vowels. Treatment targets were generally counterbalanced across equipment and order across speakers, although approximants were addressed second for all participants. Pamela and Palmer spent six sessions solely with EPG (sibilants, vowels), and three sessions solely with ultrasound (liquids); Peran and Purdy spent six sessions solely with ultrasound (vowels, sibilants) and three solely with EPG (liquids). For the final five sessions, all participants alternated within sessions between ultrasound and EPG. The difference in time allotment by equipment provided an opportunity to make a qualitative, non-experimental comparison between the two technologies. During treatment, the first and second authors (S-LPs) modeled targets and provided feedback, using speech, sign and written information. Targets changed in complexity from silent movements to isolated segments to syllables, words and finally phrases.

Prior to the everyday listener study reported here, trained listeners evaluated the pre- and post-treatment data using phonetic transcription (Bernhardt et al., 2003). The transcriptions indicated a 50% gain in consonant target accuracy post-treatment for Pamela, Purdy and Palmer, and a 28% gain for Peran. This compared with a 28% gain in vowel target accuracy for Purdy and Palmer, compared with a 16.9% gain for Pamela and a -1% regression for Peran (Bernhardt et al., 2003). Across speakers, the trained listener suggested that the most improved consonant was /r/, and the most improved vowel was /ʊ/, followed by /i/ and /u/. The question for the current study was whether everyday listeners would notice these or similar improvements.

Everyday Listener Evaluation in Study #2

Method for listener study #2.

Research assistants who had not helped with the treatment study organized the stimuli for the everyday listener study. Ten native English-speaking everyday listeners between 20 and 45 participated (6 men, 4 women). All had post-secondary education, no familiarity with disordered speech, and normal hearing as screened at 25 dB from 500 to 4000 Hz.

Nonsense word samples from the ultrasound recordings were selected for the everyday listener study, because they had also been used in the trained listener study (Bernhardt et al., 2003). The stimuli were controlled in terms of phonetic context (/h/, /ɑ/ and /b/), and thus it was assumed that the listener could focus on the test segment. The following targets were included: vowel treatment targets /hib, hɪb, hub, hʊb/, vowel observation targets /heɪb, hɛb, hɑb/, consonant treatment targets /sɑ, hɑs, ʃɑ, hɑʃ, lɑ, hɑl, rɑ, hɑr/and consonant observation targets /tʃɑ, hɑtʃ/. The audio-recorded sound files were transferred to a Macintosh computer using Adobe Premiere 6.0 and Macromedia SoundEdit 16. The sound files differed within and across speakers in terms of degree of background noise. This discrepancy was the result of different recording levels during the probes rather than different signal-to-noise ratios, because recordings were made under constant ambient noise conditions. Overall signal amplitude was reduced for the louder files using SoundEdit 16, yielding equally loud tokens across the listener sessions, and bringing relative background noise to within 3 dB for all tokens.

As was the case with the first listener study, the small *n* favored within- rather than between-speaker analyses. The stimuli were entered into PsyScope 1.2.5 (1993). However, because PsyScope 1.2.5 (1993) could not easily identify the source of individual tokens after randomization, listeners were asked to make only within-speaker judgments for the four individual speakers. Data from a fifth speaker was used for a short training session. The within-speaker rating procedure may have resulted in a practice effect for the listeners. However, speaker order was counterbalanced across listeners, and pre- and post-treatment tokens were randomized within the same sets. In order to ensure data recoverability from PsyScope, word-initial consonants, word-final consonants and vowels were also in separate sets. The five target consonants (/s, ʃ, l, r, tʃ/) yielded 10 consonant sets, that is, 5 sets of word-final consonants and 5 sets of word-initial consonants. Each consonant set contained 20 randomized tokens, 10 pre- and 10 post-treatment. Equal numbers of the five consonant syllables were presented in random order. There were seven vowel targets, and thus seven vowel sets. Each set contained 20 vowel syllable targets, 10 pre- and 10 post-treatment tokens, with all vowels represented as indicated above. Because of the 20-token limit per set, one vowel could only be tested twice in each set. However, all listeners heard all seven sets from each speaker, ensuring that listeners rated all available vowel tokens over the seven sets.

A 3-point judgment scale was adopted for the listener evaluation, following Ertmer and Maki (2000). The 3-point scale allowed listeners to provide an intermediate rating for tokens that were somewhat like the target. Listeners attended two testing sessions of about 90 minutes each. In the first session, they practised with the training set, and rated all test sets from two speakers. In the second session, they rated sets from the other two speakers. Listeners wore Koss UR-20 headphones and sat in an IAC sound booth. Instructions were given orally and on the computer screen. Listeners were instructed to focus on the target (e.g. "FINAL CONSONANT" of the word), and to ignore the rest of the word. They were told to press "1" if the stimulus sounded "EXACTLY" like the target, "2" if the stimulus sounded "SOMEWHAT" like the target, and "3" if the stimulus sounded "NOT AT ALL" like the target. They could escape from the program at any time. For consonants, the printed consonant syllable appeared in English orthography on the computer screen at the same time as the sound recording was played back (e.g., *saw, hoss, shaw*, etc.). The listener then registered his or her rating on the computer keyboard (1, 2 or 3), within 5 seconds. In pilot testing, trained listeners could not respond sufficiently quickly to the nonsense word vowel stimuli. Thus, for the vowels, a familiar word containing the target vowel was presented on the screen. The listener was asked whether the vowel heard in the auditory stimulus was the same as the one in the word on the screen: /i/ - *need*, /I/ - jig, /u/ - *rude*, /ʊ/ - *good*, /eɪ/ - *raid*, /ɛ/ - *bed*, and /ɑ/ - *log*. No other external reference was given to the listeners, in order to ensure that they would use their own internal reference without experimenter biasing.

With PsyScope 1.2.5, responses occasionally did not register because of the speed of a response. Unfortunately, if a given stimuli set had a missing response, the program could not identify which item was missing, and thus all data from that set had to be eliminated. The number of incomplete consonant sets was just slightly greater than chance (5.8%), but 21-25% of vowel sets across speakers had missing sets. There was no difference in terms of missing response sets across speakers, and no individual speaker's vowel data had to be eliminated altogether. There was a bimodal difference in listener responses; five listeners had missing responses in less than three data sets, and the other five had missing responses in five to seven data sets. The listener groups did not differ significantly in *rating* proportions ('1,' '2,' and '3' responses), however, showing that speed of response was probably the affecting variable. If listeners had more than one missing response set for a given speaker and consonant or vowel category, all data from that speaker were eliminated for that listener and category as a cautionary measure. Otherwise, complete data sets were pooled across listeners within speakers, giving a final number of listener tokens as follows: Peran: Vowels (V) — 840; Consonants (C) — 1720; Palmer: V — 920; C — 1920; Purdy: V — 1000; C — 1890; Pamela: V — 1040; C — 1900.

Results and discussion for listener study #2.

Results were evaluated within speaker, with Table 3 showing overall consonant and vowel ratings. Across listeners, average pre-treatment consonant ratings showed fairly similar ratings in the low on-target range for Purdy, Peran and Pamela: 1.61 to 1.76. For Palmer, the average rating was 2.16, that is, in the intermediate accuracy range. A Wilcoxon's Summed Ranks non-parametric test was used to test pre-post differences, because of the small sample. Post-treatment, by speaker, average consonant listener ratings improved significantly for Palmer and Pamela, to 1.52 and 1.62 respectively ($p < .01$). Purdy showed a slight non-significant improvement, and Peran a slight non-significant regression. The most improved consonant ratings across speakers were for /r/, with individual speaker variability for the sibilants and /l/. Average ratings for vowels across listeners pre-treatment were in the mid on-target range for all speakers, from 1.4 for Pamela to 1.67 for Palmer. Post-treatment, Purdy and Palmer showed a small but significant increase in ratings to 1.42 and 1.50 respectively ($p < .01$). Pamela and Peran showed a small, non-significant improvement. The vowel /i/ showed the most improved listener ratings.

Space precludes a detailed comparison with the trained listener study (Bernhardt et al., 2003) in terms of speaker ratings. However, as predicted, there was general congruence. Palmer was rated most severely pre-treatment. Palmer, Pamela and Purdy all showed greater gains in consonant ratings than Peran, and Purdy and Palmer showed greater gains in vowel ratings than the other two. The everyday listeners also agreed with the trained listeners in rating /r/ as the most improved consonant across speakers. Rankings for less-improved consonants differed; the everyday listeners rated the sibilants overall more highly than did the trained listeners, perhaps showing a greater tolerance for dentalization of sibilants. Trained and everyday listener evaluations disagreed on the most improved vowel. According to the everyday listener ratings, the most improved vowel was /i/, whereas in the trained listener study, /ʊ/ was evaluated as most improved, followed by /i/. The /ʊ/ vowel may be difficult for everyday listeners to evaluate, partly because of English orthography, where the "oo" can be /u/ or /ʊ/ or because it is a lax vowel and therefore less salient. The rating task was generally problematic for the vowels, as attested by the number of abandoned vowel data sets.

The listeners' differential response to the various speakers further confirms that future research will have to consider speaker profiles. The everyday listener rating

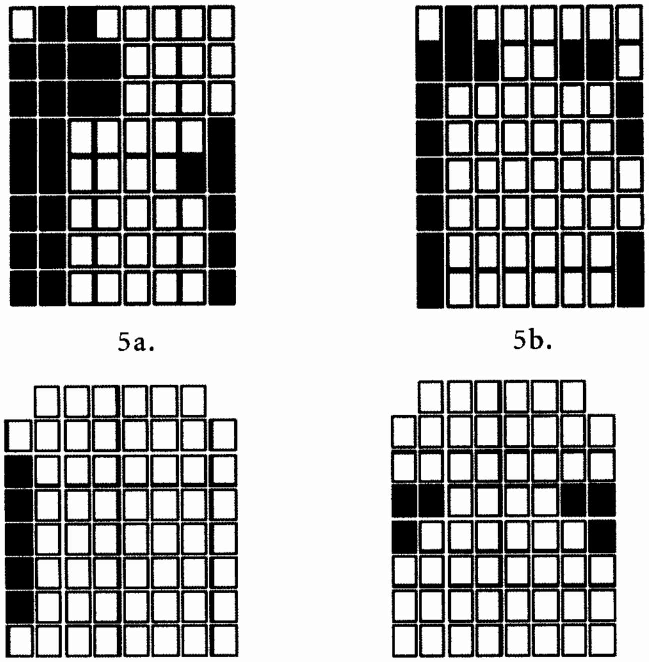
Table 3

Everyday listener ratings (1,2,3) of speakers' pre- and post-treatment consonants and vowels

C or V	Speaker	Average rating pre-Tx	Average rating post-Tx
Consonants	Purdy	1.61 (0.77)	1.55 (0.72)
	Peran	1.69 (0.73)	1.70 (0.75)
	Pamela	1.76 (0.81)	1.62* (0.72)
	Palmer	2.16 (0.81)	1.52* (0.69)
Vowels	Pamela	1.40 (0.63)	1.38 (0.58)
	Purdy	1.53 (0.69)	1.42* (0.62)
	Peran	1.57 (0.69)	1.51 (0.68)
	Palmer	1.67 (0.73)	1.50* (0.69)
C	All	1.81 (0.81)	1.59* (0.72)
V	All	1.53 (.69)	1.45* (.65)

Note: Standard deviation in parentheses. "1" rating: "exactly like the target;" "2" rating: "somewhat like the target;" "3" rating: "not at all like the target." Based on ratings of 10 everyday listeners of the consonants /r/, /l/, /s/, /ʃ/, /tʃ/, and vowels /i/, /ɪ/, /u/, /ʊ/, /ɑ/, /ɛ/, /eɪ/.
*p < .01 on Wilcoxon's Signed Ranks tests.

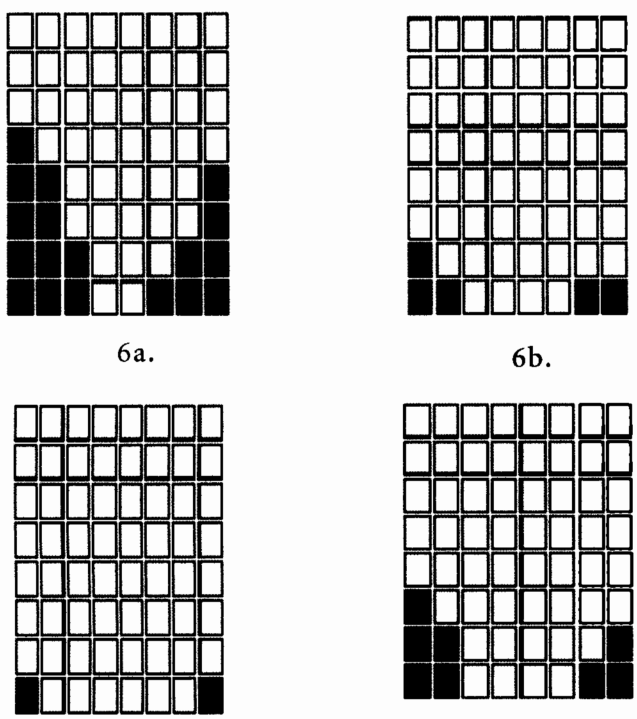
Figure 5



5a. 5b. 5c. 5d.

Figure 5a-5d. EPG pre- and post-treatment images for Pamela's /s/ (5a, 5b pre-post) and /r/ (5c, 5d pre-post). Post-treatment, /s/ has a narrower groove, and /r/ more symmetrical contact.

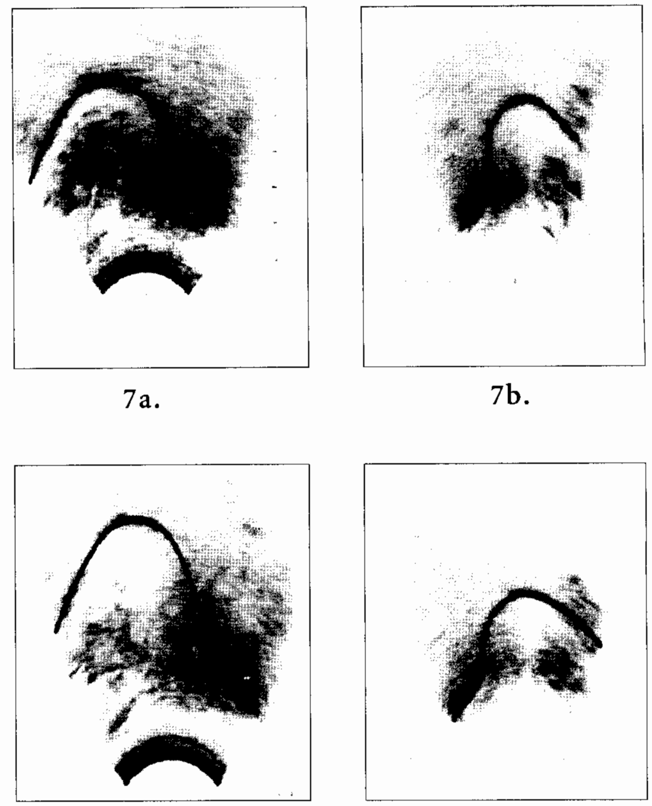
Figure 6



6a. 6b. 6c. 6d.

Figure 6a-d. EPG images for Peran's vowels /ʊ/ (6a,6b pre-post) and /u/ (6c, 6d pre-post). The /u/ is advanced, and the /ʊ/ retracted after treatment.

Figure 7



7a. 7b. 7c. 7d.

Figure 7a-d. Ultrasound images for Pamela's /i/ (7a, 7b pre-post treatment) and /ɪ/ (7c, 7d pre-post). The /i/ is advanced and heightened in comparison with /ɪ/ post-treatment.

scale did discriminate between speakers and speech targets. Vowel ratings were problematic, however, in this study. Furthermore, individual listener references for the "1," "2" and "3" ratings remain unknown. If a speaker has an average 1.5 rating on some target, it is not clear what that implies for the target or everyday communication. The rating scale provided less specific observations than the word identification task as used in the first study.

As a final note on the second study and as a further tutorial on ultrasound and EPG, sample pre- and post-treatment images are given in figures 5-7. Figure 5 shows a narrowing of the groove for /s/ post-treatment for Pamela (figures 5a, 5b), and a more symmetrical contact pattern for /r/ (figures 5c, 5d). The EPG images for Peran show the /ʊ/ tongue-palate contact moving backward (figures 6a, 6b) and the /u/ contact moving forward (figure 6c, 6d). The ultrasound images for Pamela's /i/ (figures 7a, 7b) and /ɪ/ (figures 7c, 7d) show a differentiation post-treatment, with the /i/ showing a higher tongue body and advanced tongue root. Overall, these images correspond generally to the listener perspectives. For further discussion of perceptual-articulatory convergence in ultrasound studies, see Adler-Bock (2004).

Conclusion

The current paper had two purposes: (1) to provide an introduction to EPG and ultrasound as tools in speech habilitation, and (2) to present preliminary everyday listener observations concerning pre- and post-treatment samples collected for two exploratory studies at UBC, as a foundation for future research and clinical initiatives.

Comparing EPG and ultrasound, they provide different types of dynamic information about the tongue during speech production. EPG shows tongue-palate contact patterns from the tongue tip to the back of the hard palate for mid and high vowels and lingual consonants. Ultrasound images tongue shape, location and configuration for all vowels and lingual consonants. In practical terms, EPG is much less expensive than ultrasound, and WIN-EPG has built-in analysis capabilities. However, ultrasound does not require individualized artificial palates for each client, and strides are being made in data quantification (see the aforementioned websites). In terms of speech habilitation, both tools appear to hold promise. Research is needed to compare the relative benefits of each in speech habilitation, and their merit in comparison with other technologies and methods across a variety of speakers.

In working towards clinical implementation of EPG or ultrasound, it is important to know whether the benefits outweigh the costs. The ultimate test of any treatment methodology is through randomized clinical trials, with large numbers of participants, comparison and control groups, blinding of S-LPs to conditions and data, rigorous baseline and treatment protocols, multiple types of evaluations and control of external variables. As a prelude to such trials, additional small *n* studies may provide further information as to the relative merit of the technologies. The everyday listener studies reported in the current paper raise a number of questions for future research in speech habilitation with visual feedback technology, particularly in terms of speaker profiles and evaluation methods. More research is needed to determine the relationship of speaker variables to outcomes, for example, age, etiology or severity. In terms of evaluation methods, a number of issues were raised by the everyday listener studies. In the second study, everyday listeners agreed with a previous study by trained listeners on the most improved speakers and consonant, with near agreement on the most improved vowel. Although this congruence is encouraging, one aspect of the everyday listener data suggests that it may be important to consider the relative impact of the degree of change in future studies. Everyday listener observations across the two studies showed minimal pre-post differences for 7 of 11 speakers. This may be because the everyday listeners rated six of the pre-treatment samples relatively highly, leaving minimal room for improvement. Trained listeners might have noted minor improvements through narrow phonetic transcription or instrumental analyses. However, it is not known what the impact of small differences might be in speakers' everyday conversations. More research is needed to compare everyday and trained listener observations and to determine the impact of different types and degrees of change on conversational intelligibility. In the studies reported here, everyday listener measures of intelligibility were used, specifically word identification and accuracy judgments. Word identification may be more ecologically valid than accuracy judgments, because conversation involves word identification. However, accuracy judgments can differentiate between speakers and samples. Thus, both can contribute information to the evaluation process. Although everyday listener observations bring the 'real world' into the evaluation process, future research also needs to bring the interaction between speaker and listener into focus through comprehensibility studies (Visser, 2004; Yorkston et al., 1996). Qualitative studies gathering the viewpoints of the speakers and their conversation partners may also be illuminating, and are currently underway in the UBC research program.

In terms of clinical application, S-LPs currently have limited or no access to EPG or ultrasound for their clients unless they can form partnerships with university centres or hospitals engaged in research. Cleftnet UK has been established and may also be a future possibility in Canada. Meanwhile, the information about phonetics and treatment evaluation presented in this paper may provide S-LPs with some new ideas for daily clinical practice and its evaluation.

Acknowledgements

Thank you to the following: the speakers and listeners; S-LPs David Loyst and Shannon Muir, Study #1; research assistants Dana Bryer, Dana Kanwischer, Jonathan Howell and Marie Jette; Dr. Katherine Pichora-Fuller of the University of Toronto and Dr. Linda Rammage of the Vancouver Hospital Voice Clinic, committee members for Rhea Williams' Master's thesis; and the reviewers of this paper. For funding, we thank the Canadian Foundation for Innovation, the BC Medical Services Foundation, UBC's Hampton and HSS funds and the Variety Club.

References

Adler-Bock, M. (2004). Visual feedback from ultrasound in remediation of persistent /r/ errors: Case studies of two adolescents. Unpublished Master's thesis, University of British Columbia.

Bernhardt, B., Fuller, K., Loyst, D., & Williams, R. (2000). Speech production outcomes before and after palatometry for a child with a cochlear implant. *Journal of the Association of Rehabilitative Audiology, 23,* 11-37.

Bernhardt, B., Gick, B., Adler-Bock, M., & Bacsfalvi, P. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics and Phonetics, 19,* 605-617.

Bernhardt, B., Gick, B., Bacsfalvi, P., & Ashdown, J. (2003). Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics and Phonetics, 17,* 199-216.

Bernthal, J., & Bankson, N. (2004). *Articulation and phonological disorders,* Fifth Edition. Boston, MA: Allyn and Bacon.

Black, H. (1999). An evaluation of palatometry in the treatment of /r/ for a hard of hearing adolescent. Master of Science graduating essay, University of British Columbia.

Bridges, C.C. & Huckabee, R.M. (1970). A new visual speech display- its use in speech therapy. *Volta Review, 72,* 112-115.

Dagenais, P. (1992). Speech training with glossometry and palatometry with profoundly hearing impaired children. *The Volta Review, 94,* 261-282.

Delattre, P. C. & Freeman, D. C. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics, 44,* 29-68.

Derry, K., & Bernhardt, B. (2000). Palatometry intervention in relation to body (structure and function), activity, and participation. Poster presented at the VIIIth meeting of the International Clinical Phonetics and Linguistics Association, Edinburgh, Scotland.

Ertmer, D., & Maki, J. (2000). Comparison of speech training methods with deaf adolescents: spectrographic versus noninstrumental instruction. *Journal of Speech, Language & Hearing Research, 43,* 1509-1523.

Fletcher, S., Dagenais P., & Critz-Crosby, P. (1991). Teaching consonants to profoundly hearing-impaired speakers using palatometry. *Journal of Speech and Hearing Research, 34,* 929-942.

Frattali, C. (1998). *Measuring outcomes in speech-language pathology.* New York, NY: Thieme.

Gibbon, F., Crampin, L., Hardcastle, W., Nairn, M., Razzell, R., Harvey, L., & Reynolds, B. (1998). Cleftnet (Scotland): A network for the treatment of cleft palate speech using EPG. *International Journal of Language & Communication Disorders, 33,* supplement, 44-49.

Gibbon, F., Hardcastle, W., Dent, H., & Nixon, F. (1996). Types of deviant sibilant production in a group of school-aged children, and their response to treatment using electropalatography. In M.J. Ball & M. Duckworth (Eds.), *Advances in clinical phonetics* (pp. 115-149). Amsterdam: John Benjamins.

Gibbon, F., & Wood, S. (2003). Using electropalatography (EPG) to diagnose and treat articulation disorders associated with mild cerebral palsy: A case study. *Clinical Linguistics and Phonetics, 17,* 365-374.

Gick, B. (2002). The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association, 32,* 113-122.

Hardcastle, W., Gibbon, F., & Scobbie, J. (1995). Phonetic and phonological aspects of English affricate production in children with speech disorders. *Phonetica, 52,* 242-250.

Hardcastle, W., Jones, W., Knight, C., Trudgeon, A., & Calder, G. (1989). New developments in EPG. A state of the art report. *Clinical Linguistics and Phonetics, 3,* 1-38.

Kent, R., Miolo, G., & Bloedel, S. (1994). The intelligibility of children's speech: A review of evaluation procedures. *American Journal of Speech-Language Pathology, 3,* 81-95.

Klasner, E., & Yorkston, K. (1999). Dysarthria in ALS: A method for obtaining the everyday listener's perception. *Journal of Medical Speech-Language Pathology, 8,* 261-264.

McLeod, S., Roberts, A., & Sita, J. (2003). The difference between /s/ and /z/: More than +/- voice? Poster presented at the American Speech and Hearing Association Convention, Chicago, IL, November, 2003.

Maki, J. (1983). Application of the speech spectrographic display in developing articulatory skills in hearing-impaired adults. In I. Hochberg, H. Levitt, & M. Osberger (Eds.), *Speech of the hearing impaired: Research, training and personnel preparation* (pp. 297-312). Baltimore: University Park.

Michi, K. Yamashita, Y., Imai, S., & Yoshida, H. (1993). Role of visual feedback treatment for defective /s/ sounds in patients with cleft palate. *Journal of Speech and Hearing Research, 26,* 277-285.

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R. Kent (Ed.), *Intelligibility in speech disorders: Theory, measurement and management* (pp. 11-34). Amsterdam: John Benjamins.

Schuster, L., Ruscello, D., & Toth, A. (1995). The use of visual feedback to elicit correct /r/. *American Journal of Speech-Language Pathology, 4,* 37-44.

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics, 21,* 291-311.

Stone, M. (2005). A guide to analyzing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics, 19-,* 455-501.

Stone, M., & Lundberg, L. (1996). Three-dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America,* 99(6), 1-10.

Volin, R. (1991). Micro-computer based systems providing biofeedback of voice and speech production. *Topics in Language Disorders, 11,* 65-79.

Yorkston, K., Strand, E., & Kennedy, M. (1996). Comprehensibility of dysarthric speech: Implications for assessment and treatment planning. *American Journal of Speech-Language Pathology, 5,* 55-66.

Williams, R. (1998). Outcomes of palatometry therapy as perceived by untrained listeners. Unpublished Master's thesis, University of British Columbia, 1998.

World Health Organization. (2001). *ICF: International classification of functioning, disability and health.* Geneva: World Health Organization.

Author Note

Please direct enquiries to Barbara Bernhardt, School of Audiology and Speech Sciences, University of British Columbia, 5804 Fairview Avenue, Vancouver,BC V6T 1Z3, bernharb@interchange.ubc.ca

Received: November 16, 2004
Accepted: October 4, 2005