Critical Review: The effectiveness of ultrasound technology as a visual biofeedback tool on the productive speech intelligibility of adolescents and young adults with a hearing impairment

Lauren Gallagher
M.Cl.Sc (SLP) Candidate
University of Western Ontario: School of Communication Sciences and Disorders

This critical review examines the effects of ultrasound technology as a visual biofeedback tool as an adjunct to speech therapy on the productive speech intelligibility of adolescents and young adults with moderate to profound hearing impairments. Study designs include: case study and single subject design. Overall, research suggests a facilitative role of this technology in developing more typical tongue placement for challenging phonemes particular to this population to increase intelligibility, with hope for long-term effects. However, the evidence base for this field of research is limited and the studies reviewed demonstrate methodological weaknesses. Recommendations for future research and clinical practice are provided.
.

Introduction

Adolescents and young adults with a hearing impairment often present with unintelligible speech as a result of residual articulatory errors despite being aided and despite receiving speech therapy. Due to reduced auditory feedback for certain sounds, those with a hearing loss may not have adequately received the perceptual information necessary to distinguish classes of sounds, with a resulting production that is delayed or disordered (Ruffin-Simon, 1983). Some of the characteristic errors include: omission of lower intensity speech sounds, substitution of stops for fricatives/and or lack of clear contrast between stops and fricatives and stops and nasals, overuse of front consonants, inappropriate voicing, and neutralized vowels (Smith, 1975; Tobey, Pancamo, Staller, Brimacombe & Beiter, 1991; Dagenais, 1992; Tye-Murray, Spencer & Woodworth, 1995).

Speech therapy for articulation has been traditionally based on the motor theory of speech perception, which teaches the articulatory gestures to sounds. It has been successful with visible speech sounds, but it is difficult to use this technique for sounds produced within the oral cavity (Ertmer et al., 1996). In addition, the feedback from this training can be subjective. Visual feedback training based on acoustic measures with the use of a spectrograph has been used as an alternative to this method with success, meeting both the demand of supplementing the auditory information with visual feedback and using objective measures (Ertmer et al., 1996). Visual feedback has also been met with success with Electropalatography (EPG) (Dagenais et al., 1994).

The use of ultrasound as a visual biofeedback tool in speech therapy for this population is a relatively new approach. It offers a dynamic, real time visual display of the articulatory gestures of the tongue. It is non-invasive, and does not rely on the use of a transducer or any other material inside the mouth to measure tongue movement, unlike palatometry, glossometry and EPG. A transducer is placed against the underside of the chin, above the larynx, and images are created when high-frequency ultrasound waves reflect as they pass from tissue into air just above the tongue's surface. The tongue dorsum, tip, and root are visible using the sagittal images, and additional information on tongue grooving is available with coronal images, though the palate is not visible and must be added to the image separately as a reference (Stone, 2005). This technology therefore allows otherwise invisible gestures of sounds to be imaged in real time as an adjunct to speech therapy for individuals who have experienced reduced auditory feedback due to a hearing loss.

Objectives

The primary objective of this paper is to critically evaluate existing literature regarding the use of ultrasound technology in improving the productive speech intelligibility of adolescents and young adults with moderate to profound hearing impairments. The secondary objective is to propose evidence-based practice recommendations for clinical application and future research.

Methods

Search Strategy

Computerized databases including PubMed and Scopus were searched. The following key terms were targeted: ultrasound AND (hearing-impaired) OR (hard of hearing) OR deaf AND (speech-therapy) OR (speech intelligibility). Examination of reference lists from retrieved articles revealed further studies for review. The search was limited to studies involving adolescents and young adults.

Selection Criteria

Studies were chosen for inclusion in this critical review based on the following criteria: (a) a study of ultrasound treatment was completed; (b) baseline and post intervention results were reported in the single subject designs; (c) speech production skills were the variable of focus; (d) subjects were aided with hearing aids or cochlear implants; and (e) conclusions and recommendations for the use of ultrasound were provided.

Data Collection

Results of the literature yielded five articles that matched the aforementioned criteria: a case study and four single subject designs.

Results

Foss, Whitehead, B., Paterson, & Whitehead, R. (1990) explored the practical use of ultrasound in a pilot case study design. Two subjects with an unidentified hearing impairment with bilateral aids were recruited, but only results from one subject are discussed as the second subject dropped out. The subject participated in two 45 minute sessions per week over the course of 10 weeks, half of which consisted of real-time feedback from the ultrasound in conjunction with speech-therapy and the other half traditional speech therapy. Pre and post intervention assessments were conducted. Authors report an improved ability to mark transitions between vowels, improved placement of target phonemes, and a reduction of oral-nasal confusions.

The study provides helpful information on the tolerance of the subject to this type of technology, noting that he was readily able to integrate feedback from his own tactile awareness of tongue position with the visual images and real-time scanner. There are several major weaknesses to the study, however, which make it difficult to draw conclusions. Neither the type or severity of hearing loss is indicated, no pre or post data are presented for analysis, and the same speech pathologist administered and evaluated the assessment tests, resulting in equivocal evidence for this study.

Bernhardt, Gick, Bacsfalvi & Ashdown (2003) evaluated the effect of ultrasound and EPG on the speech habilitation of four adolescents age 16-18 with moderate-severe hearing loss and moderately unintelligible speech in a single subject design study. All participants had received many years of speech therapy prior to the study, but had not learned several of the target contrasts or phonemes. Subjects participated in two treatment blocks of therapy: 9 weekly sessions and 5 weekly sessions, using both technologies. In the first block, one pair of subjects had six weeks of ultrasound and 3 of EPG, and the opposite for the other pair. All subjects received both technologies for the last block. All participants worked on the same targets: coronal fricatives, liquids, and tense-lax vowel contrasts. Transcribed speech samples were completed by trained listeners pre and post and compared to adult targets: Percent Consonants Match (PCM) and Percent Vowels Match (PVM). Authors reported significant improvements in speech production across all participants. There was significant improvement in treatment targets over non-treatment targets, and the greatest gains were made on consonants that were absent or marginal pre-treatment. There was no advantage found of one technology over the other.

A strength of this study is the randomization to visual feedback technology (EPG and ultrasound), enabling a comparison of one over the other. This study demonstrated weakness in its analysis of target accuracy in that it relied solely on a transcription of a speech sample consisting of a carrier phrase followed by a word containing the target sound and compared to adult targets. Furthermore, agreement between the transcriptions of the two trained listeners was 63% for the phoneme targets, and 84% for direction of change in accuracy between pre and post samples. Authors stated the direction of change to be the most important component; however, reliability of that measurement is called into question given a potential bias of the transcribers being the authors of the study. Despite these methodological weaknesses, the rapidity with which the subjects seemed to acquire the targets they struggled with in years of therapy point to a promising facilitative role of the technology with more controlled testing. Evidence for this study is suggestive.

Bacsfalvi, Bernhardt & Gick (2007) evaluated the effect of EPG and ultrasound with vowel remediation as the focus in a single subject design with three 18-

year old participants with severe hearing impairments who were aided. These students also participated in the Bernhardt et al. (2003) study above and were therefore familiar with the technology. Intervention took place twice a week for six weeks. Each session began with an awareness component, followed by demonstration, then practice with still and moving images with either ultrasound or EPG separately. Outcome measures were: phonetic transcription, acoustic analysis of vowel formant values and EPG tongue-palate contact patterns. Authors conclude there were notable changes observed for all vowels across all the speakers, most notably /i/, reported by authors to be difficult for this population due to high second and third formants . Subject 1 showed change on all vowels in some dimension, Subject 2 showed change across all vowels in some dimension, with /i/ showing improvement on all measures, and subject 3 showed positive change across all vowels. In general, authors conclude 8 of the 15 vowels showed gains, but it is unclear under which of the three measures these gains are reported.

A strength of this study was using several measures to determine accuracy of target production: transcription, acoustic analysis, and EPG contact patterns, an improvement over the 2003 study which looked solely at transcription. In addition, raters outside of the study conducted both the Praat acoustic analysis as well as the EPG analysis. Authors of the study performed the transcription. However, it was rare that all three measures agreed with one another for each vowel. Some vowels showed improvement on transcription, but no change or unfavourable change on another measure. It is difficult, therefore, to determine whether or not these changes were significant. One reason for this inconsistency may relate to the data from the age-matched controls. The hearing impaired subjects were compared to hearing speakers from the area, but data from only one male and one female were used. It is possible that a data set averaged from a large collection of control speakers would have been more valid and discriminating. In terms of the clinical questions posed for this review, it is not possible to attribute these gains to one technology over the other, but rather to the summed effect of both EPG and ultrasound as visual feedback technology. For this reason especially, evidence for this study in terms of ultrasound alone is equivocal.

In a single subject, non-concurrent multiple baseline design, Bacsfalvi (2010) investigated the effect of ultrasound technology on the attainment of the lingual components of /r/ in three adolescents with severe-profound hearing loss with cochlear implants. All subjects had received prior speech therapy, and two of the subjects did not produce any of the components of /r/ before the intervention. Intervention took place weekly for an unspecified number of weeks. The lingual components of /r/ as identified in the study are: tongue root retraction, tongue tip retroflexion/curling or tongue blade bunching and tongue midline grooving. Each component was learned before the next was introduced. Outcome measurements were ultrasound image analysis and listener evaluations consisting of a yes/no response for rhotic quality. The three listeners were speech-pathologists who had previous experience with ultrasound imaging and clinical research with North American /r/. All participants learned the gestural components of /r/ with ultrasound, and one was able to accurately produce it in isolation and at the word level.

Due to the inherent challenge in acquiring and teaching /r/, it is noteworthy that all participants made gains in the gestural components of this phoneme in a matter of weeks. Ultrasound imaging allows otherwise invisible aspects of /r/ production to be seen, namely tongue root retraction, tongue tip curling and tongue midline grooving. An independent evaluator of the ultrasound imaging who was blind to the chronological order of the images as well as the identity of the speakers, provided reliability of the author's judgment with an agreement of 95% for all three participants. In a study of /r/ remediation with ultrasound for hearing speakers (Adler-Bock et al. 2007) authors used acoustic analysis of formant values to judge particularly F3 values. /r/ is characteristically produced with a lower F3 value and small difference between F2 and F3 values (Delattre & Freeman, 1968; Guenther et al., 1999; Westbury et al., 1998). Adler-Bock et al. (2007) suggest that this change in F3 value is achieved with constriction of the pharynx, palate and lips. A detailed F3 acoustic analysis may have been a helpful addition to ascertain the proper positioning of the articulators, especially as the yes/no judgments for rhotic quality are highly subjective. In this study, listener agreements were quite low, ranging from 63% - 75%, though apparently acceptable when judging speech production for the hearing-impaired population. The evidence in this study suggests a beneficial impact of ultrasound as a visual biofeedback tool.

Bacsfalvi & Bernhardt (2011) conducted a follow-up study to determine the long-term outcomes 2-4 years post speech therapy for adolescents with ultrasound and EPG. Subjects had participated in the Bacsfalvi, 2010 /r/ study as well as the Bernhardt et al. (2003) study. A within-subjects design was used. Subjects were administered the *Computerized Articulation and Phonology Evaluation System* (CAPES). No results are offered, however. Target phonemes were recorded from speakers in words or phrases. Subjects 1-3 were targeting /r/ and subjects 4-7 were targeting /r/, /s/, /ʃ/ and /i/. Expertly trained listeners identified each phoneme by choosing a region representing a scale of accuracy for that phoneme on a computer screen. Data provided by the listeners for consonants were arranged into categories of percent accuracy. Vowels were matched against gender matched formant values. Inter-rater reliability ranged from 62% - 81% These perceptual judgments by the listeners showed that five out of seven speakers either maintained their level of performance or continued to generalize post treatment.

As the authors note, there were a number of limitations/weaknesses to this study: a different recording device and location were used, which may have affected the acoustic signals; only perceptual analyses were used, whereas additional measures were used in previous studies; there was no comparison to normative data for consonants; and the vowels did not have normative data for region. Additionally, some of the participants had continued to receive speech therapy, whereas others had decided to enter into the Deaf community and use ASL. Given these weaknesses, it is difficult to make a conclusive statement about the long-term effects of this technology. Certainly, this study suggests it may be possible to maintain lasting effects, but further research must be conducted to confirm.

Discussion

As a group, these studies are similar in that they are based on small sample sizes, either as a single case study or single subject designs with fewer than ten participants. Though these small samples cannot adequately represent the population, they yield critical information of treatment efficacy for each individual involved in the study. These small studies, if done well, can serve to guide the clinician in his/her decisions regarding treatment.

In general, results appear to support a facilitative role of ultrasound as an adjunct to speech therapy for this population. In order to draw conclusions about the efficacy of this tool, there are some points to consider concerning the methodology of these studies. There may have been an element of bias in several of these studies where the authors performed the transcriptions. As well, several of the studies lacked normative data with which to compare the subjects' productions. One of the studies did not provide baseline or post-intervention results (Foss et al., 1990), and in general, agreement between raters on quality of productions was questionable.
Across these studies, a range of measurements were used to determine the subjects' change in production: Bernhardt et al. (2003) used transcription only, while Bacsfalvi et al. (2007) looked at acoustic analysis, EPG contact patterns, and transcription. Bacsfalvi (2010) used perceptual measures of yes/no and ultrasound image analysis, and Bacsfalvi & Bernhardt (2011) used perceptual measures for phoneme accuracy, compared with some normative data for vowels, but not consonants. Foss et al. (1990) did not report measures. In addition, in the Bacsfalvi et al. (2007) study, while there was widespread improvement on target sounds across participants, not all the measures agreed with one another, some reporting gains while others not. This prompts a question into what the best measurement of improved production is, and whether it is expected that improvements in transcription will always align with improvements in acoustic analysis. Given the disparate measurements and inconsistency between measures of the same production, the evidence must be interpreted with caution.

Another point that merits consideration is the fact that four of the studies were conducted in the same Laboratory, with the same author involved in all four. While this is not necessarily problematic, it is possible there may have been bias in subject recruitment or choice of target phonemes.

More research is required to confirm these questions in order to arrive at a definitive conclusion, but the gains made by participants, particularly in the quick acquisition of the gestures of /r/ compared to years of therapy, as well as gains made by all participants in /i/, are promising, especially as visual biofeedback has been shown to be beneficial to this population.

Conclusion

Ultrasound as a visual biofeedback tool has tremendous potential benefit for individuals with a hearing loss with its real-time display of otherwise invisible articulatory gestures of the tongue, offering a cue to supplement impaired auditory perception. As it is a comparatively expensive tool, the purchase of an ultrasound may not be justified given the limited evidence, but with continued research, it could be a highly useful tool in addition to traditional speech therapy.

Recommendations

Future research should be conducted with the following points in mind:

- Larger sample sizes could be used in order to better reflect the population. To enhance the evidence, a study that includes a control group receiving traditional speech therapy without a visual aid could be used to compare with those receiving ultrasound as visual feedback.
- Studies should be conducted with another group of researches, recruiting novel participants
- Reliable outcome measures should be established to accurately reflect the changes in production, and these should be employed consistently in order to compare across studies. Specifically, there should be an inclusion of naive listener perceptual ratings, as well as self-rated and family-rated scores.
- Studies should be longer in duration in order to allow for the use of target phonemes in words, sentences and conversation, as intelligibility in conversational speech is the ultimate goal.

Clinical Implications

With more research, Speech-Language Pathologists could incorporate ultrasound imaging into traditional therapy to provide hearing impaired clients with visual biofeedback of tongue motions in order to help them achieve more typical patterns. This would be especially useful for learning new articulatory patterns (Bernhardt et al., 2003), especially those that are otherwise invisible. It has the potential of being educational, interactive, and motivating. Use of this technology also has the potential of reducing therapy time from years to months (Bacsfalvi, 2012), thereby being cost effective. In addition, it could act as a diagnostic tool to help clinicians discover incorrect patterns, such as lack of constriction or incorrect tongue tip placement to guide intervention. Stone (2005) details how to set up and position equipment to get the best resolution as well as how to interpret images, which would be a useful guide for beginning clinicians.

References

Adler-Bock, M., Bernhardt, B., Gick, B., & Bacsfalvi, P. (2007). The use of ultrasound in remediation of English /r/ in two adolescents. *American Journal of Speech Language Pathology*, 16(2), 128-139.

Bacsfalvi, P. (2010). Attaining the lingual components of /r/ with ultrasound for three adolescents with cochlear implants. *Canadian Journal of Speech Language Pathology and Audiology*, 34, 206-217.

Bacsfalvi, P. (2012). Ultrasound and Speech Habilitation [slideshow]. Retrieved from http://www.caslpa.ca/english/events/2012_handouts/Ultrasound%20_An_essential_tool_for_Speech_Language_2012%20CASLPA%20Participants-%20handout%20format.pdf

Bacsfalvi, P., & Bernhardt, B. (2011). Long-term outcomes of speech therapy for seven adolescents with visual feedback technologies: Ultrasound and electropalatography. *Clinical Linguistics & Phonetics*, 25 (11-12), 1034-1043.

Bacsfalvi, P., Bernhardt, B., & Gick, B. (2007). Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. *Advances in Speech-Language Pathology*, 9(1), 36-45.

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 & Phonetics*, 17 (3), 199-216.

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

Dagenais, P.A., Citz-Crosby, P., Fletcher, S.G., & McCutcheon, M.J. (1994). Comparing abilities of children with profound hearing impairments to learn consonants using electropalatography or traditional aural-oral techniques. *Journal of speech and hearing research*, 37, 687-699.

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

Ertmer, D.J., Stark, R.E. & Karlan, G.R. (1996). Real- time spectrographic displays in vowel production training with children who have profound hearing loss. *American Journal of Speech-Language Pathology*, 5(4), 4-16.

Foss, M., Whitehead, B., Paterson, M. & Whitehead, R. (1990). Ultrasound as a visual feedback aid for the hearing-impaired. *Journal of Diagnostic Medical Sonography*, 2, 80-86.

Guenther, F., Espy-Wilson, C., Boyce, S., Matthies, M., Ruffin-Simon,C. (1983). Relationship between production and perception of speech: Some developmental issues. In I. Hochberg, H. Levitt & M.J. Osberger (Eds.). *Speech of the Hearing Impaired: Research, Training, and Personnel Preparation*. (pp.233-250). Baltimore: University Park Press

Smith, C., (1975). Residual hearing and speech production in deaf children. *Journal of Speech and Hearing Research*, 18, 795-811.

Stone, M. (2005). A guide to analyzing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics*. 19(6-7), 455-502.

Tye-Murray, N., Spencer, L. & Woodworth, G. (1995). Acquisition of speech by children who have prolonged cochlear implant experience. *Journal of Speech and Hearing Research, 38, 327-337.*

Westbury, J., Hashi, M., & Lindstrom, M. (1998). Differences among speakers in lingual articulation for American English /r/. *Speech Communication*, 26, 203–226.

Zandipour, M., & Perkell, J. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, 105, 2854–2865.