Oral and oropharyngeal cancer: speech outcomes and the need to monitor change over time. Phillip Doyle

Oral and Oropharyngeal Cancer: Speech Outcomes and the Need to Monitor Change Over Time

Philip C. Doyle, Ph.D.

Voice Production and Perception Laboratory, Department of Otolaryngology and Reconstructive Surgery, Schulich School of Medicine and Dentistry, Doctoral Program in Rehabilitation Sciences, and Communication Sciences and Disorders, University of Western Ontario, London, ON, Canada

pdoyle@uwo.ca

Abstract

Objectives: Despite the long history of surgery as the primary treatment for cancer of the oropharyngeal structures, a relative paucity of information on speech outcomes currently exists. While the importance of this voice and speech performance is clearly identified as an essential outcome along with chewing and swallowing, there remains inconsistency in how outcomes for speech are obtained and interpreted. It is clear that surgical treatment for oral and oropharyngeal tumors and subsequent reconstruction has a direct influence on the integrity of the human vocal tract. As a result, surgery and reconstruction holds the significant potential for alteration of signal transmission characteristics of the vocal tract. This presentation provides a brief summary of existing data on outcomes and serves to raise several questions related to a need for research. Further, preliminary data obtained in our laboratory using established methodology to assess the acoustic structure of English vowels in individuals who have undergone surgery for oral cancer is presented for consideration.

Patients/Materials and Methods: Following a cursory review of the literature, results from a preliminary study of 11 participants (3 women, 8 men) who had received surgery for treatment of oral cancer are presented. All participants had undergone surgery for oral tumors and received subsequent reconstruction using a radial free forearm flap. A group of age- and gender-matched control participants were also studied. All participants produced multiple productions of 10 English vowels consistent with methods outlined by Peterson and Barney (1952); formant frequencies were then determined via spectrographic analysis.

Results: Data obtained revealed remarkable similarity across the speaker groups for values obtained for formants 1, 2, and 3 (F1, F2, and F3). When proportional relationships between F1 and F2 were compared, no differences between surgical and control participants were identified. Overall, and in the presence of substantial surgical reconstruction, these 11 surgical participants
appear capable of modifying their vocal tracts so as to adequately code the basic and essential acoustic requirements for the 10 vowels studied.

Conclusions: These preliminary findings demonstrate that traditional and fundamental methods of vocal tract assessment, namely determination of formant frequencies through acoustic analysis, can serve as a valuable index of postsurgical vocal tract function and signal transmission. These acoustic data represent one specific measure of vocal tract transmission characteristics that could be exploited in additional analyses of the postsurgical system. The lack of significant differences between these groups does not, however, suggest that the speaker groups are equivalent with respect to overall speech production capabilities. The current data provide an initial step toward establishing standardized methods of post-treatment speech evaluation.

Key words: oral cancer, oropharyngeal cancer, vocal tract, speech acoustics, speech rehabilitation

1. Introduction

The treatment of oral and oropharyngeal malignancies poses numerous and often considerable challenges that cross multiple areas of concern. As with all malignant conditions and the associated surgical treatment options, aspects of oncologic safety are paramount. Medical management will focus on eliminating disease, but intricately tied to the concern of oncologic safety comes a desire to retain functional capacity (Teichgraeber, Bowman, & Goepfert, 1986) and optimize “quality of life” (Doyle, 1994). When considering the functional capacity of the upper airway, primary considerations have typically centered on aspects of verbal communication, deglutition, and swallowing; beyond issues related to survival, these three functional domains are widely agreed to comprise facets of critical importance to outcomes in those treated for oral and oropharyngeal cancers (Logemann et al., 1993; Matthews & Lampe, 2005). These areas also are linked inextricably to those broad domains of functioning that are most frequently addressed in relation to QOL, namely, myriad facets of physical, psychological, and social capacity (Doyle, 1994; Myers, 2005; Rieger, Dickson, Lemire, et al., 2006). Yet the ability to speak, as well as to chew and swallow without difficulty, remain at the forefront of any effort to document the success of cancer treatment regardless of modality (Urade et al, 1987).

The determination of a successful or unsuccessful outcome in those diagnosed with oral and oropharyngeal malignancies must always be conceptualized in a multidimensional fashion. However, for the purpose of this paper, the information presented is directly focused on the distinct area of speech capacity in the post-treatment period. The emphasis on verbal communication in this paper should not be interpreted to imply that deglutition and swallowing (or other areas of communication functioning) are of less importance—they are not; collectively, these functions are critical elements of outcome measurement that seeks to provide evidence for treatment effectiveness and success and its impact
on the individual. Nevertheless, the dynamic processes associated with the generation of voice and speech and ultimately, the ability to systematically quantify and monitor the speech signal, require careful attention. Thus, this rather focused topic will be addressed exclusively in the remainder of this paper. Consequently, several key areas will be identified in an effort to provide a framework for the documentation of potential post-treatment changes, as well as a means of evaluating ongoing changes in such deficits. Perhaps the most important issue of concern in the context of speech outcomes is directly related to inconsistent and variable methods of evaluating outcomes. As a result, “evidence” becomes inconsistent both in its presentation and interpretation.

Dylan et al. (2007) reviewed 12 articles that reported some level of speech outcome, but identified that there “…were few studies that included precise objective outcome measures.” (p. 3). One issue observed by these authors is that some oropharyngeal resections are complicated by not only their inherent surgical variability, but may also create additional functional complexities when laryngectomy and/or surgical-prosthetic voice restoration (tracheoesophageal [TE] puncture) is involved. It is important to note, however, that despite these complexities, many studies have ultimately applied a practical approach to assessment that is directed toward one’s global communication activities for daily tasks such as phone use. Though broad, such approaches may offer more refined, individual profiles of post-treatment speech (communication) outcomes as a marker of evidence.

In order to facilitate the discussion to follow, a cursory overview of several issues is necessary. This will initially involve a discussion of: (1) oral and oropharyngeal structures and the composite structures and functions of the vocal tract, (2) surgical intervention for oral cancer and the subsequent reconstruction of surgical defects that result from tumor excision, and finally, (3) the ability to monitor changes to the vocal tract acoustically in the context of post-treatment follow-up. Throughout these sections, a selective review of information from the literature will be provided. Although the information included in this review is not in any manner exhaustive, information provided is judged to provide a framework for data presented later in the paper. Next, and as a direct outgrowth of this work, some suggestions for future research will be offered for consideration. Finally, although references to specific reconstruction methods may be noted along the way, it is beyond the scope of this written document to address the differences in those surgical techniques in detail at this time.

1.1. Cancer and the Oral and Pharyngeal Cavities

The oral and oropharyngeal cavities are comprised of numerous, highly integrated anatomical structures which lie superior to the inlet of the voice generation mechanism of the larynx. Consequently, this important region exhibits substantial complexity both anatomically and physiologically. Basic structures include the lip, alveolus, teeth, floor of mouth, buccal tissue, palate, velum, tongue, and pharynx, as well as the region of the tonsillar bed, base of tongue, maxilla, and mandible. Collectively, these structures ultimately comprise the
vocal tract, a variable length tube (cavity) that exists from vocal folds to lips (Kent, 1993; Kent & Read, 1992). The vocal tract is a dynamic region that varies considerably in response to linguistic requirements and the acoustic representations dictated by those requirements for the purpose of verbal communication. Because of the critical nature of the vocal tract on speech production, alterations in the vocal tract hold considerable importance relative to subsequent alterations in the acoustic speech signal that represents the end product of speech. More specifically, abnormalities in the structural integrity of the vocal tract, concurrent changes in its physiologic function, and compensatory strategies that may be employed to delimit such changes must be considered in a multidimensional fashion (Morrish, 1988). However, in regard to any efforts to evaluate speech outcomes in this heterogenous population, it is apparent that standards have not been clearly defined. One critical shortcoming of much of the work conducted in the area of speech outcomes in those treated for oral and oropharyngeal cancers relates to the fact that no standard approach or protocol appears to be used for speech assessment. While a number of studies in the literature do indeed use sound, rigorous methodology and a rational approach to gathering and providing outcome measures on speech, considerable differences do exist. Additionally, terminology used and the descriptors employed are also problematic. In fact, Lorenz and Alam (2003) have identified a tendency for speech assessments to be based on somewhat vague definitions such as “…all had speech intelligible enough for communication” (p. 234). Interestingly, these authors also note past concerns in the literature that suggest the importance of speech “quality” on measure of speech intelligibility proper. That is, while intelligibility may be the ultimate target or a given assessment (Kent, Weismer, Kent, & Rosenbek, 1989), and in fact is the true end product of speech communication, changes in the overall “quality” of the voice/speech signal may provide a substantial confound to judgments made and/or their interpretation.

While many methods employed for measuring speech outcomes are indeed reasonable and sound from a procedural and methodological point of view, the inherent degree of variability in methods from study-to-study is considerable. Consequently, the ability to reasonably and validly compare across literature contributions is often quite limited. In essence, the variability in tumors treated, differences in the amount of tissue resected, the amount and type of reconstruction utilized, multiple modalities of treatment, concerns of medical complications and associated treatment morbidities, as well as numerous other factors makes such comparison daunting at best. Work by Clark et al (2006) has offered a detailed evaluation of operative morbidity associated with pharyngeal reconstruction in an attempt to identify and describe early and late complications. The data obtained provided predictive data on complications and morbidity in relation to the extent of surgical defect, type of reconstruction, and method of initial treatment. These data would appear to provide valuable information which suggests that speech outcomes might be better understood if one can understand more about the details of surgery and associated complications;
differences are indeed substantial from individual to individual—no two surgical procedures are identical.

Uwiera, Seikaly, Rieger, Chau, and Harris (2004) reported somewhat paradoxical information on those undergoing hemiglossectomy; while a significant difference was observed between pre- and postoperative intelligibility measures at the single-word level, no differences were identified for the sentence materials. These data confirm the relative limitations of using particular speech materials in measuring post-treatment speech outcomes. Minimally Uwiera et al. provide support for using multiple measures of intelligibility (Lariviere, Seilo, & Dimick, 1975). This finding raises additional questions concerning the need for developing standard methods of intelligibility assessment in this population. In the absence of standardized approaches, development of a credible evidence base will be restricted and our ultimate ability to accurately detect outcome will be restricted. With this caveat provided, the information to follow may provide an initial snapshot of more obvious concerns and areas of research need in association with treatment of oral and oropharyngeal cancers.

1.2. The Vocal Tract

When seeking to understand the functional integrity of the vocal tract, many factors come into play. The vocal tract functions as a dynamic region that in large part permits the uniqueness of the speech signal for each individual. It is the combined relationship between the larynx and the vocal tract that makes each speaker’s verbal communication unique. In the context of oral and oropharyngeal cancers, two additional considerations emerge. Changes in the vocal tract always hold the potential to alter the uniqueness of one’s speech communication. First, the general structural integrity of the vocal tract must always be considered when addressing cancer treatment for tumors within the oropharyngeal system. Second, one also must consider the ability of the postsurgical system to grossly and finely function so as to allow the vast complexity of adjustments. While adaptation and compensation is common and not unexpected, restrictions will exist. Changes to the physiologic capacity of the vocal tract have a direct influence on the rapid and highly coordinated movements necessary for intelligible speech. Degradation in the overall function of the vocal tract creates numerous obstacles that will influence the character of speech produced.

Concerns about what defines “acceptable” speech outcomes has been raised by Butler and Lewin (2004). The primary issue pertains to the fact that in some instances, intelligibility may be defined very broadly, ranging from somewhat vague descriptions to those that are much more elaborate and operationally

1 Although the present commentary specifically addresses the functional capacity of the vocal tract secondary to surgical management of cancer, these concerns are of equal importance to other treatment modalities such as radiation and chemoradiation. All methods of cancer treatment hold the potential to disrupt normal structure and function of the vocal tract. Consequently, the methods described herein are equally applicable to other treatment populations.
defined indices of change. For example, according to Butler and Lewin, intelligibility can be judged according to whether it was “understandable by the medical personnel without the need for interpretation by the family, multiple requests for repetition, or written clarification of what was said.” (p. 500-501). Although this definition does provide a means of broadly classifying intelligibility and serves as a generalized index of communication ability, it may lend itself to varied interpretation and, hence, create problems in objectively indexing speech performance. Loss of such definitions creates the possibility for the inaccurately interpretation of existing data, and perhaps more importantly, its comparison to new data.

In regard to the problems encountered in accurately quantifying one’s post-treatment speech capabilities, some reports have provided rather contradictory information relative to speech “intelligibility” (Pigno & Funk, 2003). More specifically, in their case report, Pigno and Funk note that their patient’s speech had improved, being “intelligible but noticeably in error.” (p. 121). Though this information does support some generalized index of the level of communicative function specific to speech, a description of the “errors” noted may hold substantial importance in understanding the nature of deficits that occur and the potential for rehabilitation to improve speech intelligibility. Small deficits may create substantial problems even in the presence of compensation. Similarly, some studies have adopted use of a global 5-category rating system for classifying general intelligibility (Chow, Chan, Chow, Fung, & Lam, 2007). While categorical or cluster ratings do hold some generic assessment value, it does not always characterize the collective entities that comprise speech and how they merge together as a coherent unit. Further, the net effect of such changes to the ear of the listener are essentially unknown. Thus, fine details as well as “gestalt” measures must be considered in the context of generating evidence of outcome in those with oral and oropharyngeal cancers. Doing so can merge the speaker and the listener as a dynamic, dyadic unit specific to communication effectiveness.

Consequently, evaluation of the functioning of both the vocal tract proper and the end-product of speech as judged by the listener in the post-treatment period is essential. Doing so may serve as an objective means of determining component deficits that may then correlate with reductions in gross speech intelligibility and its component parts. One must acknowledge, however, that changes in the function of the vocal tract may involve considerable adaptation over time (perhaps even over several years) and there appears to be little if any data on this topic. Single session observations are unlikely to provide evidence that will withstand scrutiny relative to external validity. As a result, it is imperative that efforts to document acoustic, and ideally auditory-perceptual correlates of acoustic change in those treated for oral and oropharyngeal malignancies, be performed over reasonable periods of time; measurement over time is absolutely essential. Doing so will facilitate our ability to understand ongoing changes that are the result of the loss of anatomic tissue and the degree and extent of the
surgical defect(s), approaches to reconstruction, and longer-term changes due to healing, reduction in tissue swelling, concurrent influences of combined treatments, residual influences of treatment (e.g., the formation of scar tissue), and the individual’s ability to adapt and compensate for dynamic speech restrictions that occur (Doyle & Keith, 2005). Yet discrepancies in how intelligibility is defined and how such definitions overlay to one’s functional capacity persist and continue to create challenges in the interpretation of our evidence. For example, in regard to a group of participants who had been treated for aggressive orofacial cancer with submental flap reconstruction, Chow et al (2007) noted that “…all retained normal conversational function, except for two who were able to communicate with intelligible speech. (p. 500). Statements such as this, coupled with limitations in procedural standards for measuring outcomes, add further challenges to fully understanding the subsequent functional outcomes in this diverse population.

Changes within the vocal tract exist in numerous planes and ultimately, the key issues of how the vocal tract works to create specific acoustic structures necessary for “normal” speech are found primarily in the length of the vocal tract and the varied cross-sectional areas over that length. In essence, during speech the vocal tract is continually (and rapidly) being modified in shape via varied degrees of constriction as a result of tongue positioning, jaw movement, and lip shape (e.g., rounding). The net effect of these changes create varied “cavities”, each of which permits further acoustic modification of the fundamental sound source that emanates from oscillation of the vocal folds. Changes may be viewed from the perspective of simple acoustics (e.g., extent of frequency change), to more dynamic and global areas of functioning that are reflected in measures of speech intelligibility, to the composite characteristics that define each speaker.

Haughey, Taylor, and Fuller (2002) evaluated 43 subjects who had undergone resection and fasciocutaneous free flap reconstruction for cancer of the tongue and floor of mouth. Thirty of these individuals received reconstruction of the oral tongue while 13 had reconstructions of the tongue base. Overall, they reported that speech intelligibility scores were significantly better for those who received tongue base reconstructions (98% intelligibility) when compared to those with reconstruction to the oral-tongue (76% intelligibility). This work also suggested that 20% of participants experienced either medical or surgical complications. Assessment of speech intelligibility was derived from averaged scores based on a 50-item word list with judgments made by five naïve listeners. From these data, substantial variability was noted with scores ranging from 20-100%. Results revealed a mean intelligibility score of 98% for those undergoing tongue base procedures with a mean of 68% for those with oral tongue reconstructions. This amount of variability has been reported previously in the literature relative to anterior tongue defects in association with relatively low mean intelligibility scores (Bokhari, & Wang, 2007; Rentschler & Mann, 1980; Leonard, 1983; Leonard & Gillis, 1990; Mah et al., 1996; Pauloski et al., 1993, 1998).
Chien, Su, Hwang, et al. (2006) presented data for a group of 20 participants who received total glossectomy for base of tongue cancers and 7 who had subtotal glossectomy for tumors of the oral tongue. The reconstruction method of choice for these participants was either a radial forearm free flap (RFFF) or anterolateral thigh flap. Based on their data, Chien et al. indicated that approximately 90% demonstrated “intelligible” speech postoperatively. Similarly, Biglioli, Liviero, Frigerio, et al. (2006) reported information on a group of individuals who had undergone partial glossectomy and RFFF reconstruction as a means of preserving nerve function for the tongue. Overall, increased satisfaction was reported by those receiving neurofasciocutaneous (sensate flaps) procedures when compared to those undergoing non-sensate reconstruction methods. In fact, the level of satisfaction was almost double for those in the sensate group (78% vs. 43%). The sensate flap group also was found to exhibit higher “articulation” scores via objective evaluation; however, only 44% of participants in that group demonstrated such results, thus raising numerous questions about the methods of assessment and the associated influence of the extent and degree of tumor excision and reconstruction. Further, questions related to the potential interactions between speech performance and satisfaction can be reasonable raised and beg the question of determining if speech performance and satisfaction are indeed linked.

Seikely et al. (2003) reported findings from 27 participants who were diagnosed and treated for oropharyngeal cancer. Within this group, 18 provided complete data for both speech and swallowing outcomes. Speech data were obtained prior to surgery, at one month before radiotherapy was initiated, and at some point between 6-9 months postsurgery and following the completion of radiation. The results of this work suggest that the amount of tissue resected for base of tongue tumors did not have any significant influence on either speech or swallowing. However, from the standpoint of speech acoustics and aerodynamics, Seikaly et al. recommend that additional study is warranted for those that undergo resection of the soft palate of half or greater. Although post-treatment intelligibility measures were good, the nature of stimuli used was deemed to be of importance. Once again, differential findings for word intelligibility and sentence intelligibility were noted; this raises questions related to the content of the stimuli used for assessment of speech intelligibility as it is an essential factor to consider in determining evidence-based outcomes. Finally, these authors did note the importance of “drop out” of participants on the data obtained. The concern of participant drop out rates have been raised previously in the clinical literature (Colangelo, Logemann, Rademaker et al., 1999) and such considerations may carry substantial value in assessing the range of true outcomes in multiple areas of functioning following treatment. If individuals cannot be monitored when at their worst, can the data generated be validly interpreted?

Determining meaningful measures offers a considerable challenge in the clinical population who undergo treatment for oral and oropharyngeal cancers and no
tried and true approach is universally accepted. Therefore, our research group in 
the Voice Production and Perception Laboratory (VPPL) at the University of 
Western Ontario has initiated a preliminary approach in efforts to assist our 
understanding of the postsurgical speech production system with an emphasis on 
vocal tract transmission characteristics. Although we are keenly interested in 
the end-product of speech and its interaction with the listener as part of the auditory-
perceptual process, our efforts to date have primarily focused on quantitative 
evaluation and analysis of the acoustic signal at the vowel level of performance. 
The work to be presented herein has been directed toward efforts to objectively 
document the influence of surgical treatment on the production of English vowels. 
As noted earlier in this paper, our focus on vowel production does not 
discount the importance of other areas including the larger concern of one’s overall level 
of speech intelligibility, and/or general aspects of communicative effectiveness as 
a critical means of social interaction, and of course, the larger dimension of 
quality of life. Yet our interest in vowel generation provides what our group 
believes is an important and fundamental index of postsurgical functional speech 
capacity for those undergoing surgery for cancers of the oral region.

In the work to be described, it was anticipated that intrinsic acoustic properties of 
a vowel (i.e., formant frequencies) might be degraded due to the extent of the 
surgical resection, tissue reconstruction, or development of scar tissue and/or 
decreased mobility of the tongue or other moveable structures within the vocal 
tract. These factors also are likely to influence one’s overall speech intelligibility 
and have in fact been documented for many years (Skelly, 1973). Several 
studies have reported that speech intelligibility in those who had undergone 
partial glossectomy was inversely proportional to the amount of tissue removed 
and the subsequent adequacy of the tongue-palate valve (Bokhari & Wang, 
2007; Dios et al, 1994; Michi et al, 1989; Michiwaki et al, 1990; Soutar & 
McGregor, 1986; and others).

Changes in one’s ability to successfully “valve” regions within the vocal tract in 
the postsurgical period have been a fruitful area of study. In order to optimize 
reconstruction following total glossectomy in an effort to reduce speech deficits, 
Sanger, et al (2000) modified the forearm flap method of reconstruction to include 
use of a portion of the brachioradialis muscle. Their decision to augment the 
basic flap procedure was driven by observations that simple flaps could reliably 
resurface tongue defects, but they did not appear to adequately permit palate 
and tongue relationships necessary for efficient speech. Thus, the use of the 
brachioradialis was intended to decrease deficits associated with posterior bulk in 
the vocal tract. Through this alteration in the flap procedure, it was anticipated 
that speech would be improved, potentially in relation to aerodynamic parameters 
and oral-nasal relationships for speech.

Gerden et al. (2003) have provided data that is quite comprehensive relative to 
speech. These researchers evaluated speech capacity via use of several 
methods of assessment for oronasal relationships for speech, along with self-
perceptions of speech using a questionnaire format. This work was carefully conducted and confirmed that issues of nasality (alterations in the nasal resonance of the speech signal) must be considered in those undergoing palatomaxillary resections with reconstruction. The importance of such work has been addressed in comprehensive fashion by Leeper, Gratton, Lapointe, and Armstrong (2005). However, various reconstructive surgeries have emerged as a means of restoring articulatory function, and in fact, seek to enable the vocal tract to mimic a somewhat normal system. Such changes, even if minor, hold the distinct and real possibility of creating dynamic deficits in the transmission pathway with a direct bearing on the acoustic characteristics of vowels, as well as other speech sounds. In a follow-up study, Gerden, Wallace, Okay, and Urken (2004) reported findings on 12 individuals who had undergone RFFF reconstruction for defects of the hard palate. Additionally, 8 individuals who had been fitted with an obturator also were evaluated. Data gathered indicated that those who had undergone flap reconstruction reported higher satisfaction for speech, comfort, and social interactions. However, individualized details of performance were not reported. In summary, it would appear clear that the variability are the data and that efforts to describe individual performance in greater detail would be of substantial value in understanding outcomes (Muldowney, Cohen, Porto, & Maisel, 1987).

It is important at this point to specify that our concerns relative to the potential acoustic changes are derived from two, overlapping questions secondary to surgical treatment for oropharyngeal cancer. First, the obvious issue of structural change due to the extent of tissue ablation always must be considered. The clinical and experimental literatures attest to the highly variable changes that exist as a consequence of surgery in the oropharyngeal region and its subsequent reconstruction. It is clear that no two surgical procedures are identical in this population. Further, despite occasional inferences of “similarity” for a given procedure, both the short- and long-term outcomes across many dimensions of “oral” functioning will be characterized by considerable variance; surgical groups are not homogenous. This variability also carries with it opportunity for idiosyncratic behavioral changes that may either positively augment oral function for speech (and perhaps other oral functions), or detract from it. However, we have been particularly concerned with the influence of reconstruction of extensive surgical defects, particularly those that comprise substantial surface areas in the oral and oropharyngeal regions. In such instances, our interest has been further expanded in situations where a relatively large free flap may be used to reconstruct the surgical defect that exists following tumor excision. Thus, the shape, size, configuration and interaction with other structures (e.g., the tongue) may be altered substantially. Based on these factors and an underlying desire to monitor how the postsurgical oral system and vocal tract functions under the demands of verbal communication, it has been the intent of our group to approach the problem in a stepwise manner that utilizes established, but relatively simple acoustic measures as evidence for describing the functional integrity of the postsurgical vocal tract.
The information to follow represents one small, and in fact what some might justifiably argue is a rather narrow, reductionist approach to understanding this population relative to speech. But we view it as a starting point and as such, this initial phase of the project has served the needs of a larger and longer-term ongoing program of research. The application of acoustic measurements certainly offer us the ability to systematically monitor changes in the postsurgical vocal tract under relative static conditions (i.e., during vowel production). Yet beyond acoustic measures, and although not covered in this discussion, we have also attempted to look ahead and consider the important area of auditory-perceptual evaluation as part of generating evidence on the success of flap reconstruction procedures. This approach would appear to offer considerable advantages; it not only has translational value in that it allows for longstanding and well-accepted “bench research” methods to be viewed in a clearly applied fashion, but more importantly, this work may ultimately provide a straightforward means for linking acoustics and the perceptual consequences of surgical treatment for cancers of the oral and oropharyngeal regions.

1.3. Monitoring and Indexing Change Secondary to Cancer Surgery

Postsurgical alteration of the oral and oropharyngeal cavities may be judged perceptually and/or by quantifying the acoustic character of the vocal tract via a number of acoustic measures. These acoustic measures cross frequency, intensity, and temporal domains. Acoustical measures of the composition and representation of energy concentrations in the resonated vocal signal allow for the direct analysis of characteristics of the vocal and speech signal (Baken, 1987; Kent, 1979; Kent, 1993; Kent & Read, 1992; Peterson & Barney, 1952). This type of assessment was initially reported in the seminal work reported by Peterson and Barney (1952). Peterson and Barney presented speakers with a list of 10 monosyllabic words each beginning with the glottal phoneme /h/ and ending with the voiced alveolar stop consonant /d/; the interconsonantal phoneme was a vowel. The variation in this structured production task was created by altering the vowel component within the /h_d/ stimulus structure (e.g., /hI_d/, /hanD/, /head/, etc.). From these stimuli, Peterson and Barney (1952) asked groups of adult men and women and children to produce stimulus tokens. Productions were then acoustically analyzed to determine the resonant frequencies of the vocal tract (i.e., formant frequencies) in association with the different vowels produced. The underlying premise of this concept is found in the fact that the tongue works to change the general cavity shape(s) and cross-sectional area of the vocal tract. It should be noted at this point, however, that two additional physiologic features are also critical to vocal tract shape and the specific resonance characteristics associated with any vowel, namely, lip-rounding, and the excursion of mandibular movement. Collectively, cooperative and highly integrated adjustments of the tongue, lips, and jaw permit the vocal tract to be modified into a variety of shapes for the purpose of speech production (Kent, 1993; Kent & Read, 1992).
Using Peterson and Barney’s (1953) method, the purpose of the present study was to quantitatively examine acoustic structure directly related to changes in vocal tract transmission (i.e., transfer function). The potential for changes in acoustic transmission would appear to be heightened following intraoral reconstruction via RFFF reconstruction in individuals who have undergone surgery for oral cancer (Meyerson, Johnson, & Weitzman, 1980). It was hypothesized that acoustic changes in postoperative speech would be directly related to surgical changes in the vocal tract. It was anticipated that vowel production would, at least to some degree, be degraded acoustically due to the extent of the resection, reconstruction, the development of scar tissue and decreased mobility of the tongue within the vocal tract. However, various reconstructive surgeries have been used as a means of restoring articulation via increased tongue mobility. Such changes should have direct bearing on the acoustic characteristics of vowels.

Secondary reconstructive surgeries that have been evaluated in the literature include the RFFF, microvascular tongue reconstruction, and tongue mobilizations to name a few (Salibian et al, 1993). Based on several reports in the literature, secondary reconstructive surgeries are suggested to be successful in increasing speech intelligibility (Dios et al, 1994; Michi et al, 1989; Michiwaki et al, 1990). Similarly, however, Fletcher (1988) reported that in spite of reports of individuals with partial glossectomy having “reasonably intelligible’ speech” (p.232), quantitative data (e.g., spectrographic analysis) that specifically defines altered articulatory parameters and patterns of articulation are lacking. Consequently, the current study evaluated quantitative acoustic measures of vowel production in a small, selected sample of individuals who had undergone surgical resection for oral cancer and subsequent reconstruction of the surgical defect. As such, these data may be used to ascertain changes following RFFF surgery on the resonant qualities of the oral cavity for vowel production. The specific questions addressed were: (1) What are the descriptive acoustic characteristics of F1, F2, and F3 for vowel productions by individuals with intraoral resections who have undergone reconstruction via the RFFF?, and (2) Do differences exist in the proportional relationship between formants (F1:F2) for vowels produced by experimental participants when compared to age-and gender-matched normal control participants?

2. Patients/Materials and Methods

Participants for this phase of a continuing project included 11 adults (3 females and 8 males (age 50-75 years, M = 59). All had undergone oral tumor excision and reconstructive surgery using the RFFF; marginal or segmental resection of the mandible was not an exclusionary criterion. Potential participants were excluded from consideration if they had undergone more than a single surgery for oral cancer, had received pre- or postoperative radiotherapy, or exhibited tumor recurrence. Eleven age- (+/- 3 years) and gender-matched normal control participants were also identified and evaluated. All control and experimental
participants met the following criteria for inclusion: no reported history of speech problems, head or neck surgery, hearing impairments that would affect speech production, or upper respiratory infection at the time of testing.

2.1. Procedures and Stimuli

Recordings were obtained in a sound-treated audiometric suite on research quality instrumentation. Each participant was asked to produce 3 trials each of all 10 vowels according to methods described by Peterson and Barney (1952). These stimuli were part of a larger recording protocol, but all were collected at the start of the experimental recording for all participants.

2.2. Data Analysis

The current data analysis centered solely on the inter-consonantal vowel contained in the stimuli. Initially, however, a listener assessment was carried out in order to confirm that vowel productions for each of the speakers who had undergone surgery for oral cancer were representative of the intended target. Forty samples of vowel productions were randomly selected to determine if the sample produced was representative of the intended vowel. Random selection was performed in order to control for fatigue and/or practice effects related to repeated productions of each vowel. From this set of 40 vowels, one naïve listener was asked to identify the vowel by selecting it from a closed response set of all 10 target words (e.g., hid, had, head, etc.); these items were listed in normal English orthographics and the listener circled his choice following a single presentation. Listener responses were used to establish the acceptability of vowels for later acoustic analysis. The listener correctly identified 90% (36/40) of the samples presented. Of the four errors, two were attributed to a single participant (SOC 4), a 58 year old male with resection of the right lateral floor of mouth, segmental resection of the body of the mandible, and right radical neck dissection with a RFFF. It is worth noting that this gentleman presented with a slight, although clearly identifiable British accent that may have influenced errors in perception on the part of the listener. Overall, the high level of correct identification confirmed that the intended target vowels were produced appropriately and that acoustic analysis could be pursued. Based on the extent of reconstruction, we had anticipated that more of the samples would be incorrectly identified as part of the representativeness assessment. However, these results confirm that acoustic analyses could shed further light on formant structure and vocal tract acoustics.

2.2.1. Acoustic analysis

Acoustic measures of formant frequencies (F1, F2, F3) were obtained on all vowel samples for all participants in both groups. The acoustic analysis
techniques utilized are well established in the scientific literature (Kent, 1979; Kent & Read, 1992). Measures gathered were obtained using CSPEECH 4.0 (Milenkovic & Read, 1994). Initially, each stimulus (e.g., “hid”) was digitized at 20 kHz. Using a spectrographic subroutine of CSPEECH, the vowel segment of the stimuli was identified and segmented using cursors. The experimenter then listened to the segment to insure that only the vowel was present in the segment. Once confirmation occurred, an automatic acoustic analysis and formant tracking routine was initially performed on a broadband spectrogram. In instances where automatic tracking of the formants was not believed to be accurate based on visual evaluation (i.e., the visual identification of clear, distinct energy bands on the spectrographic display), the procedure was carried out by hand, with measures being taken from the midpoint of the vowel segment where energy bands were noted. Frequency values for F1, F2, and F3 were then gathered and recorded for each vowel.

The dependent variables, F1, F2, and F3, were derived by analyzing all three samples of each vowel produced by participants who had undergone tumor resection and reconstructive RFFF surgery of the oral cavity and their age- and gender-matched speakers who had not undergone surgery. Thus, a total of 30 measures (3 samples x 10 vowels) were obtained from each of the participating subjects. This resulted in the analysis of 330 samples for each group (i.e., 3 samples x 10 vowels x 11 participants). Since three formant frequency values were obtained for each vowel sample, a total of 90 values were recorded for each participant (i.e., 3 productions x 10 vowels x 3 formant measures). Once obtained, the three measures recorded for each of the formants F1, F2, and F3 were averaged for each of the 10 vowels for comparison across subjects. Using the measures obtained from analysis of F1 and F2 for each of the three samples for all 10 vowels, the respective F1:F2 proportional values were calculated and a mean ratio for each speaker (by individual vowel) was obtained. The proportional values for F1:F2 and mean values for F1 and F2 were then statistically analyzed. A total of 33 measures for each vowel were obtained (1 vowel X 3 samples X 11 speakers) with a total of 330 measures obtained for each participant group.

2.2.2. Reliability – Formant Frequencies

Approximately 15% of vowel measures for F1, F2, and F3 were randomly selected and re-measured for reliability assessments. These samples were re-analyzed by an independent external judge trained in the same analysis procedures. Agreement was determined using a point-by-point method (i.e., # of agreements/total # of measurements x 100). As we did not anticipate identical measures to be demonstrated given the nature of the acoustic analyses, we judged measures that were within 10Hz of the original measures to be in “identical” agreement. Based on reliability analyses, measures for 40% of the surgical group and 37% for controls resulted in identical frequency values when compared to the original data. Of the remaining samples, differences were quantified in 10Hz increments. That is, when a difference was observed between
measures of a formant frequency, the difference was quantified according to whether the second measure was within a specified Hz level of the original (e.g., 11-20 Hz difference, etc.). Using this method of quantifying the degree of reliability for surgical participants, 44% of the remaining measures (n=34) were within 20 Hz of the original measures; only 12% of these samples exceeded 51 Hz of the original measures. Therefore, inter-rater reliability of measures within 50 Hz of the original measures was 94% for the surgery group. For control participants, 76% were within 20 Hz of original measures, with 14% between 21-50 Hz; only 3 measures exceeded a difference of 51 Hz from the original. Hence, for the control group inter-rater reliability for measures within 50 Hz of the original measures also was 94%.

When reliability measures were evaluated by formant frequency, the following findings were obtained. For surgery participants reliability measures, F1 = 55%, F2 = 38%, and F3 = 30% identical measures when compared to original measures. Inter-rater reliability for measures within 50 Hz of the original measures was 94%, 94% and 89% for F1, F2, and F3, respectively. For controls, F1 = 50%, F2 = 44%, F3 = 17% identical measures compared to original measures. Inter-rater reliability for measures within 20 Hz of the original measures was 100% for F1, 100% within 30 Hz for F2, and 83% within 50 Hz for F3, respectively. Based on the combined results of this reliability assessment, it is believed that the acoustic measures obtained were gathered with good reliability across measures.

2.3. Statistical Analysis

From the acoustic measures of formant frequencies (F1, F2, and F3) obtained for each vowel, means, standard deviations, and the range were calculated. These data were maintained as individual data for the subjects in order to characterize individual variation, particularly in the oral cancer group. The mean values of F1 and F2 for each vowel produced by each speaker were then collapsed and used for statistical analysis between the two groups. From the absolute F1 and F2 values, a proportional (ratio) measure of formant structure (e.g., F1:F2) was calculated and a mean ratio for each speaker and each vowel was generated. Due to the independence of each of the vowel measures, statistical analyses were conducted using analysis of variance procedures with participant speakers serving as the independent variable and vowel measures the dependent variable. An a priori significance level of p < 0.05 was used.

3. Results

3.1. Formant Frequencies – Minima and Maxima Values

Mean measures for F1, F2, and F3 values for each vowel and speaker group (surgical vs. normal controls) are presented in Table 1 for comparison. As can be seen, surgical participants exhibited slightly greater mean frequency values
for both F1 and F2. For F1, these differences ranged from 6Hz ("head") to 37Hz ("who'd" and "heard"). For F2, frequency differences ranged from 25Hz ("Had") to 176Hz ("who'd"). For F3, control participants exhibited higher frequency values only for "had" and "who'd", although these differences were not substantial. The current data for both surgical and control participants do compare favorably with normative data reported by Peterson and Barney (1952). However, it should be noted that all data for our speakers were collapsed across gender, whereas males and females were segmented by Peterson and Barney. Despite this difference, the similarly of our data provide support for the notion that even in the presence of significant reconstruction following oral tumor excision, physiologic demands for the accurate production of vowels has been met.

3.2. F1:F2 Proportional Values

The mean proportional relationship values (F1:F2) were calculated for both participant groups for each vowel. A summary of these results are also presented in Table 1. While most speaker pairs (i.e., surgery vs. control) exhibited some discrepancies in absolute F1 and F2 resonant frequencies across all 10 vowels, the mean proportional values of F1:F2 for the two groups remained quite similar. That is, those who had undergone surgery were able to code the resonant acoustic properties of vowels as well as their controls based on the ratio data generated. This finding was evident in some comparisons of individual speaker pairs where absolute frequencies differed, but where proportionally relationships for F1 and F2 were retained. Therefore, it is evident that the speakers who had undergone surgery for oral cancer and RFFF retained the ability to produce F1:F2 proportional resonances which consistently translated into well-coded vowels. This finding is consistent with results of the representativeness assessment.

4. Discussion and Conclusions

These data, though clearly preliminary in nature, offer some early insights into vowel production in those treated for oral cancer. As noted, our observation that surgical participants were able to effectively meet expectations for formant frequencies and the proportional relationships between F1 and F2 so critical for perception was unexpected. However, one observation that was of even greater surprise was that related to acoustic measures of particular vowels. For example, our data confirm that our surgical participants were able to meet the acoustic requirements of “point vowels” (/i, a, u/). These three vowels require the greatest degree of physiological excursion, therefore, provide information on the overall integrity of oral structures. It is evident that some speakers who had undergone surgery for oral cancer were more limited than others in their ability to achieve these extreme points of constriction. However, their ability to retain general proportional relationships for F1 and F2 that are so important for accurate perceptual identification was observed.
At a very basic level of production, these surgical participants were collectively able to meet the acoustical requirements of all 10 vowels investigated. Thus, despite what we would anticipate to be highly variable and individualized postsurgical vocal tract structure, the capacity to manipulate the system to achieve production of distinct vowels has been documented. While the external validity of these findings certainly must be addressed with caution, it does appear that one’s ability to generate vowels with high levels of accuracy could foster a framework for investigating more detailed aspects of consonant production in dynamic speech stimuli. Minimally, if an acoustic index of vowel structure can be identified for a given speaker, it may serve as a valuable point of reference for acoustic measures of other speech entities (e.g., when paired with consonants) with potential application to other measures of dynamic speech events (acoustic transitions, etc.) and additional methods of acoustic analysis. The present data provide one facet of a rich opportunity for monitoring the acoustic outcomes of surgical treatment for malignancies of the oropharyngeal region. The ability to measure the acoustic changes that exist following surgery may then be compared to a well-established body of normative data that has existed for many years. The value of these acoustic measures is that they may also serve as a foundation for auditory-perceptual evaluation of postsurgical speech proficiency.

**4.1. Future Directions – Acquisition of Evidence**

Efforts in our laboratory continue to address the complex problems of assessing vocal tract dynamics and its effect on speech production and perception despite the challenges that persist. Because of complexities associated with the acquisition of evidence to document outcomes and comprehensively understand the influence of surgical treatment on vocal tract functioning, several research areas are currently under investigation by our group. While individual and often narrow pathways of exploration related to voice and speech production remain at the forefront of this work, the ability to link these objective data to psychophysical measures obtained through auditory-perceptual evaluation is essential. In adopting this combined approach, in the past year we have adopted the concept of “acoustic signature” as a means of evaluating those who receive surgery for oropharyngeal cancers (Kent, Kent, Rosenbek, Weismer, Martin, Sufit, & Brooks, 1992; Weismer, Kent, Hodge, & Martin, 1988). Although acoustic signature research to date has typically been applied to intelligibility issues at the word or sentence level, we have attempted to apply this approach at the vowel level as well. In some situations, we have attempted to obtain acoustic information in the presurgical period in order to provide direct evidence of outcome on the vocal tract secondary to treatment. This approach, however, does have clear and real limitations, as well as some potential ethical considerations. Most importantly, and because of the presence of a mass lesion and the potential influence of such a lesion on the physiological integrity of the articulatory structures and vocal tract, the “baseline” that is often obtained is itself non-normal. Hence, comparative evaluations at multiple points in the post-treatment period are referenced to a potentially abnormal standard. In spite of this limitation, the ability to track
changes in the system over time can serve to reflect positive changes that occur in response to factors directly associated with healing such as decreased edema of vocal tract tissues, increased levels of structural mobility, etc. In this vein, we believe that evidence provided through acoustic tracking at the level of vowel and beyond may provide an array of measures that provide tangible evidence to the individual who has undergone surgery and must deal with multiple deficits, most notably in the early months of recovery. In this regard, we do concur with Fletcher’s (1988) suggestion that the use of spectrographic analysis holds promise. In any case, whether simple or more elaborate measures are extracted from spectrographic representations, we also believe that use of acoustic measures provides a relatively simple method of addressing changes in the vocal tract that occur following radiation therapy and/or chemoradiation protocols.

Another area of recent exploration with those who undergo treatment for oropharyngeal malignancies has involved attempts to apply common acoustic measures in the frequency, intensity, and temporal domains in a combined manner so as to obtain information on relative contributions of particular features to the end product of speech. One of the concerns within this line of inquiry relates to changes that can be noted in the early period of recovery. For example, and not infrequently, as speech capabilities return to the individual in the postoperative period, it has been observed that vocal intensity may be reduced. It is suspected that these changes may reflect an easy and “careful” approach to using the speech production mechanism (both laryngeal and articulatory) to avoid discomfort or other related fears in the presence of healing. However, it is clear that the major impediment to outcomes in those treated for oropharyngeal cancers is likely found in the lack of a measurement standard. In this regard, it is increasingly apparent that objective, as well as subjective information is ultimately necessary to determine individual performance and capacity in this important clinical population. Finally, the desire to address such questions using the traditional group design may in fact be quite inadequate. Group designs may only serve to obscure variability in the data which is in reality the essence of the data we seek to obtain as a valid means of generating evidence on those treated for oropharyngeal cancer. Thus, we have begun to adopt single-subject experimental designs in an effort to provide more specific documentation of individual outcomes secondary to cancer treatment. In the months to come, we look forward to providing additional data that expands the application of acoustic measures to this important clinical population.

5. Acknowledgments

The author would like to thank colleagues Drs. Kevin Fung, John Yoo, and Howard Lampe from the Department of Otolaryngology and Reconstructive Surgery University of Western Ontario, Dr. Wayne Matthews from the Department of Otolaryngology at the University of Calgary, and members of the Voice Production and Perception Laboratory and Postlaryngectomy Research Laboratory at the University of Western Ontario.
6. References


Table 1. Means Values (in Hz) for F1, F2, F3, and F1:F2 Proportional Values for Control and Surgical Groups.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Group</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F1:F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>heed</td>
<td>Control M</td>
<td>300</td>
<td>2301</td>
<td>3035</td>
<td>.130</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>323</td>
<td>2413</td>
<td>3109</td>
<td>.134</td>
</tr>
<tr>
<td>hid</td>
<td>Control M</td>
<td>440</td>
<td>1910</td>
<td>2610</td>
<td>.230</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>470</td>
<td>1960</td>
<td>2680</td>
<td>.240</td>
</tr>
<tr>
<td>head</td>
<td>Control M</td>
<td>539</td>
<td>1895</td>
<td>2601</td>
<td>.284</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>545</td>
<td>1986</td>
<td>2678</td>
<td>.274</td>
</tr>
<tr>
<td>had</td>
<td>Control M</td>
<td>674</td>
<td>1735</td>
<td>2492</td>
<td>.388</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>691</td>
<td>1760</td>
<td>2489</td>
<td>.392</td>
</tr>
<tr>
<td>hod</td>
<td>Control M</td>
<td>693</td>
<td>1131</td>
<td>2451</td>
<td>.612</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>705</td>
<td>1201</td>
<td>2476</td>
<td>.587</td>
</tr>
<tr>
<td>hawed</td>
<td>Control M</td>
<td>659</td>
<td>1092</td>
<td>2412</td>
<td>.603</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>669</td>
<td>1169</td>
<td>2416</td>
<td>.572</td>
</tr>
<tr>
<td>hood</td>
<td>Control M</td>
<td>465</td>
<td>1301</td>
<td>2353</td>
<td>.357</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>478</td>
<td>1353</td>
<td>2397</td>
<td>.353</td>
</tr>
<tr>
<td>who'd</td>
<td>Control M</td>
<td>329</td>
<td>1007</td>
<td>2301</td>
<td>.326</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>366</td>
<td>1183</td>
<td>2228</td>
<td>.309</td>
</tr>
<tr>
<td>hud</td>
<td>Control M</td>
<td>610</td>
<td>1427</td>
<td>2426</td>
<td>.427</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>626</td>
<td>1486</td>
<td>2511</td>
<td>.421</td>
</tr>
<tr>
<td>heard</td>
<td>Control M</td>
<td>424</td>
<td>1407</td>
<td>1895</td>
<td>.301</td>
</tr>
<tr>
<td></td>
<td>Surgery M</td>
<td>461</td>
<td>1458</td>
<td>1937</td>
<td>.316</td>
</tr>
</tbody>
</table>
Figure 1. Radial forearm free flap (RFFF) of oral cavity.
Photograph compliments of K. Fung, M.D., Department of Otolaryngology, Head and Neck Oncology and Reconstructive Surgery, University of Western Ontario, London, ON, Canada.