<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Speech outcomes after glossectomy for tongue cancer: a critical review of the literature</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Other Contributor(s)</strong></td>
<td>University of Hong Kong.</td>
</tr>
<tr>
<td><strong>Author(s)</strong></td>
<td>Li, Chi-mei; 李智美</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>Li, C. [李智美]. (2004). Speech outcomes after glossectomy for tongue cancer: a critical review of the literature. (Thesis). University of Hong Kong, Pokfulam, Hong Kong SAR.</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2004</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/48784">http://hdl.handle.net/10722/48784</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>The author retains all proprietary rights, such as patent rights and the right to use in future works.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Speech outcomes after glossectomy for tongue cancer: a critical review of the literature.

Li Chi Mei

A dissertation submitted in partial fulfillment of the requirement of the Bachelor of Science
(Speech and Hearing Sciences), The University of Hong Kong, April 30, 2004.
Abstract

Information on speech outcomes after glossectomy is needed to decide on specific surgical procedures, to effectively counsel patients before the surgery, and to plan appropriate rehabilitation strategies. This study consisted of a literature review of 83 studies published between 1970 and 2004 which reported the speech outcomes of glossectomy. The review had three purposes: 1) to summarize findings on speech outcomes after glossectomy in patients with tongue cancer; 2) to investigate how speech has been measured in these studies; and 3) to provide future direction for research in this field. The study identified four main factors affecting the speech outcome: extent of surgery, surgical procedures, use of prosthesis, and speech therapy. The study also revealed a lack of clarity in the description of speech measures employed. In addition, concerns regarding the reliability and validity of the speech measures were raised. These problems made the findings on speech after glossectomy inconclusive.
Glossectomy is the surgery of the tongue. It is the primary intervention method for patients suffering from tongue cancer (Prince & Bailey, 1999). In general, the surgery could be classified as partial or total glossectomy depending on the extent of surgery. In partial glossectomy, the anterior third of the tongue is removed and in total glossectomy, the entire tongue, including the posterior third of the tongue is removed (Evangelisto, 1997).

The tongue is responsible for the production of a majority of vowels and consonants and its volume also helps to maintain the resonance characteristics in speech (LaBlance, Kraus, & Steckol, 1991). As the tongue is an important articulator, the removal of the tongue brings changes in articulation as well as the resonance characteristics of speech sounds.

In glossectomy, removal of the carcinoma with adequate margins in order to safe the life of the patient is the primary concern of the surgery (Prince & Bailey, 1999). However, apart from whether they could survive the cancer, one major question of the patients would be what their life would be like after the surgery. One of the main factors that contributes to their quality of life is the characteristics of their speech after the surgery (Mady, Sader, Hoole, Zimmermann, & Horch, 2003). Therefore, what the patient’s speech would be and what type of surgical procedures would minimize the impact on speech is an important question both for the pretreatment counseling and deciding appropriate surgical procedures to be used for these patients. An informed answer to that question would help the clinicians give realistic expectations to those patients. Information on rehabilitation is also important because maximizing patients’ communicative ability is a high priority in the rehabilitation process.

One of the earliest papers reporting the speech outcomes of glossectomy was written by Goldstein (1940) (as cited in Amerman & Laminack, 1974), who reported that patients with the tongue completely removed could still produce intelligible speech by moving the fascia floor of the mouth to positions in the oral cavity to simulate the movements of the tongue. Since then, there have been many reports on speech after glossectomy. There have been four main foci: the effect of glossectomy on speech; comparisons of the effect of
different surgical reconstruction procedures on speech outcomes; the effect of different prosthesis; and the effect of speech rehabilitation. All these areas are important for the benefit of patients. With knowledge in these areas, patients could anticipate their speech outcome and appropriate surgical techniques or intervention method could be used to help the patients.

As in any other field of studies, given that measurement of speech production could be affected by different variables, such as the method of measurement and the listener bias (Kent, 1996), it is important that a valid and reliable speech measurement was employed. This is especially important if the aim of the study is to compare different intervention method on the speech outcome and to attribute the changes in the speech quality to the intervention method being investigated.

In the field of glossectomy, it had been suggested by some authors that the method of measuring speech outcome was not objective or there was a lack clear description of the procedures used. For example, Amerman & Laminack (1974) found that few studies on speech after glossectomy used objective methods or described their methods in details. They also found that few studies investigated speech rehabilitation for glossectomy in depth and they urged that more research should be carried out for planning speech therapy and predicting the outcome. Prince & Bailey (1999) identified the same problem and suggested that there was a lack of objective results about functional outcomes after glossectomy. Pauloski et al. (1993) also suggested in their literature review that the speech function after glossectomy were usually described in general terms. They suggested that the studies usually involved few subjects. Inappropriate combination of different types of resection or reconstruction types and insufficient description of the reconstruction procedures in the surgery were also common problems in these researches.

The above comments were useful for the researchers to improve their reports on speech function after glossectomy. However, as these comments only came from the general impression of the authors, there is a need to identify the flaws in current speech measurement...
systematically so as to provide quantifiable data for critical analysis of the problems in speech measurement. This could be done by a critical review of the literature.

Critical reviews of speech outcomes of surgery have been undertaken in other areas of speech disorder. For example, Whitehill (2002) carried out a critical review on the assessment of speech intelligibility in the patients with cleft palate by collecting data on intelligibility measurement systematically and found a lack of report on reliability. Whitehill (2002) also found that some measures used were not valid for intelligibility. Studies such as this provide important implications and insights for further studies.

As suggested above, a general impression of the field of speech on glossectomy was that this issue has not been considered in detail. Moreover, there seems to be a lack of control of bias and lack of clarity among studies in the methods used to measure speech production after glossectomy. Given that no detailed critical review has been carried out on the assessment of speech in the field of glossectomy, it is hoped that such a critical review of the literature will not only provide a summary of knowledge to date, but also critical analysis of the methodology and procedures used so that directions for future research in the field could be determined. It is also hoped that the review could provide quantifiable data to compare with the previous review articles to reveal any similarity or differences in the findings.

This study aims to address three main questions: (1) What do we know about speech outcomes after glossectomy for tongue cancer? (2) How are we measuring speech outcome in these patients? (3) What should be the future direction for research in this field?

Method

Traditionally, there was no systematic method to conduct a review article. As explained by Jankowski & Dozier (2003), traditional review mainly depended on the reputation of the reviewer and individual bias might occur because of the personal experience of the reviewer. In order to ensure objectivity of this review, ideas were borrowed from the approach used in many of the review studies for evaluating and comparing the findings.
between clinical trials in the field of health care nowadays – the systematic review. “A systematic review is a research article that identifies relevant studies, appraises their quality and summarizes their results using a scientific methodology.” (Kahn, Kunz, Kleijnen, & Antes, 2003 p. 1). It differs from traditional review because in a systematic review, the relevant primary research and their results were collected through a standardized procedure (Glasziou, 2001). It involves a number of steps: (a) question formulation, (b) finding studies, (c) appraisal and selection of studies, (d) summary and synthesis of relevant studies, and (e) determining the applicability of results. Although the current review did not aim at comparing the findings of different clinical trials, it was thought that idea could be borrowed from the systematic review so that a scientific review could be conducted.

As a matter of fact, critical reviews on the measurement of speech outcomes in the area of cleft palate speech have been carried out by Whitehill (2002) and Lohmander & Olsson (2004). They employed methods similar to a systematic review by searching the literature and recording the result of the relevant studies systematically. This study will employ similar strategies as these two studies in conducting the literature review.

**The Search**

In order to conduct a comprehensive review of the literature, a thorough literature search was carried out to cover as many related studies as possible. The search procedure involved both electronic and hand searches. In the electronic search, two electronic databases were utilized: Ovid Medline, covering over 7,000,000 journals (1970 to 2003 December) and ISI Web of Science (1970 to 2003), covering over 29,000,000 journals. For Ovid Medline, two keywords (speech or articulation) were combined with the term glossectomy and were searched as subject headings. For ISI Web of Science, the same keywords were used in the topic search function. The search was limited to studies published in English language. In addition to electronic searches, hand search was also carried out. Relevant studies cited in the journal studies and edited books were searched for using the search engine of library of the
University of Hong Kong. A summary of the search procedures is shown in Figure 1.

Combining the results from the two databases, and from the hand search, a total of 163 studies were identified. However, 24 of these were not available from our library. Therefore, a total 140 studies were retrieved. These studies were then examined with the inclusion criteria to determine if they should be included in the analysis of data.

**Application of inclusion criteria**

From the large corpus of studies, only those investigating speech after glossectomy were included in this literature review. That is, the inclusion criteria was that speech after
glossectomy was reported in the study. Even if speech was not the main focus of the study, the study was included if it presented speech results after glossectomy. Among these studies, it was found that the speech result of the same patient was reported in different papers. These studies were also included because the methodology for speech measures involved and the presentation of results in these papers were different from each other.

The exclusion criteria included (1) studies in which the speech result in the patients with glossectomy was reported together with patients receiving other surgery so that the result of speech after glossectomy could not be isolated from other surgery, (2) studies in which glossectomy was done for open bite, macroglossia or other defects not related to oral cancer, and (3) review studies and chapters that discussed about the speech after glossectomy but did not provide empirical data. This is because the current review mainly focused on the measurement of speech outcomes in research studies. However, these articles were used as references for background information and for comparison with the result of this review to investigate any similarities or differences between the current review with previous reviews.

After applying the inclusion and exclusion criteria, a total of 83 studies were found, which reported speech after glossectomy in patients with tongue cancer. The studies were shown in the appendix. It should be noted that it is possible that not all studies reporting speech after glossectomy were retrieved by this method. However, it was thought that the current literature search method used was objective and, together with the hand search, most studies related to the current topic was included.

**Development of the coding sheet**

In order to conduct a systematic review of the literature on speech outcomes after glossectomy, all studies were coded according to different categories. A coding sheet was developed for this purpose. The coding system addressed the following questions:

1. what is the primary focus of the study? (Subcategories included effect of glossectomy on speech, effect of surgical method on speech, effect of different prostheses on speech, or effect
of speech therapy on speech); (2) what was the research design employed? (for example, single case, single group, or multiple groups); (3) what type of measurement was used to reflect speech ability (for example, perceptual analysis, acoustic analysis, physiologic analysis); (4) if a perceptual method was used, what perceptual parameters were employed; (5) how were the perceptual parameters defined and measured; (6) what speech sample was used for assessing speech production (i.e., nonsense syllable, single words, sentences, or conversation); (7) what type of listeners were involved (for example, experienced speech-language-pathologists, surgeon, or naïve listeners); (8) how many listeners were used to assess speech production? (9) was listener reliability reported (intrajudge or interjudge)? (10) was statistical analysis employed in the study? (11) what was the conclusion of the study?

**Interrater reliability of the coding**

Coding accuracy may be affected by individual differences and by error. To ensure that the coding had been conducted appropriately, an experienced researcher in the field was invited to recode ten (about 10%) of the studies, selected randomly. The interrater reliability for all the categories was above 95%. The major source of differences between the coders arose as the result of summarizing the conclusion (coding category 11). Differences were resolved following discussion between the coders.

**Results**

**Distribution by Year and Source**

Table 1 shows the number of studies by the year in which they were published. There was a significant increase of number of studies reporting speech after glossectomy beginning in the early 1980s. The 83 studies were published in 31 different journals and one book. As shown in Table 2, five journals (Archives of Otolaryngology - Head and Neck Surgery, Journal of Prosthetic Dentistry, Journal of Speech and Hearing Disorders, Laryngoscope, and Plastic and Reconstructive Surgery) were the main sources of the studies and together accounted for 37 (44.6%) studies. The remainder of the studies were distributed evenly in 25
journals and one book, with each carrying one to three studies.

Table 1. Distribution of studies by year

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of Studies</th>
<th>% (n=83)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970 – 1974</td>
<td>4</td>
<td>4.8</td>
</tr>
<tr>
<td>1975 – 1979</td>
<td>2</td>
<td>2.4</td>
</tr>
<tr>
<td>1980 – 1984</td>
<td>14</td>
<td>16.9</td>
</tr>
<tr>
<td>1985 – 1989</td>
<td>15</td>
<td>18.1</td>
</tr>
<tr>
<td>1995 – 1999</td>
<td>10</td>
<td>12.0</td>
</tr>
<tr>
<td>2000 – 2004</td>
<td>16</td>
<td>19.3</td>
</tr>
</tbody>
</table>

Table 2. Distribution of studies by source

<table>
<thead>
<tr>
<th>Source</th>
<th>No. of Studies</th>
<th>% (n=83)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Journal of Prosthetic Dentistry.</em></td>
<td>9</td>
<td>10.8</td>
</tr>
<tr>
<td><em>Archives of Otolaryngology - Head and Neck Surgery.</em></td>
<td>8</td>
<td>9.6</td>
</tr>
<tr>
<td><em>Journal of Speech and Hearing Disorders.</em></td>
<td>7</td>
<td>8.4</td>
</tr>
<tr>
<td><em>Laryngoscope.</em></td>
<td>7</td>
<td>8.4</td>
</tr>
<tr>
<td><em>Plastic and Reconstructive Surgery.</em></td>
<td>6</td>
<td>7.2</td>
</tr>
<tr>
<td><em>Head and Neck.</em></td>
<td>4</td>
<td>4.8</td>
</tr>
<tr>
<td>Others (25 journals and 1 book)</td>
<td>42</td>
<td>50.6</td>
</tr>
</tbody>
</table>

Description of the surgical procedure

When describing the area of resection in the tongue, most studies (81, 97.6%) used a general description by describing the surgical resection in general terms such as “most of the tongue was resected” or by dividing the glossectomy into partial, subtotal, or total glossectomy. Two studies described the glossectomy by using surgical mapping (Bressmann, Whitehill, Sader, Samman, & Hoole, 2002; Bressmann, Whitehill, Sader, & Samman, in press).

Primary Focus of the study

The primary focus of the studies reviewed was categorized, and is summarized in Table 3. The majority of the studies reviewed (32, 38.6%) focused on outcomes of speech after glossectomy in general. Thirty-one (37.3%) studies focused on the effect of surgical procedures on speech. Fifteen (18.1%) focused on the effect of prostheses on speech. Four (4.8%) focused on the effect of speech therapy on speech after glossectomy. One study was about the effect of tracheoesophageal puncture (TEP) on glossectomy speech (Meeker,
Lavertu & Hicks, 1997). It described a case of a patient with total laryngectomy and total glossectomy who underwent TEP and was then reported to be able to produce intelligible speech to achieve a functional level of communication.

### Table 3. Primary focus of studies reviewed

<table>
<thead>
<tr>
<th>Primary Focus</th>
<th>No. of Studies</th>
<th>% (n=83)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of glossectomy on speech</td>
<td>32</td>
<td>38.6</td>
</tr>
<tr>
<td>Effect of reconstruction procedure on speech</td>
<td>31</td>
<td>37.3</td>
</tr>
<tr>
<td>Effect of prosthesis on speech</td>
<td>15</td>
<td>18.1</td>
</tr>
<tr>
<td>Effect of speech therapy on speech</td>
<td>4</td>
<td>4.8</td>
</tr>
<tr>
<td>Others</td>
<td>1</td>
<td>1.2</td>
</tr>
</tbody>
</table>

### Research design

Information of the research design is shown in Table 4. Most studies utilized a single group design (46, 55.4%) and a single case study (21, 25.3%). Twelve (14.5%) investigated two groups and four (4.8%) studies utilized a multiple group design with three to four groups. Among the studies that involved two or more groups of subjects, only one was a randomized independent group study, where subjects were randomly assigned to different groups (Skelly, Donaldson, Fust, & Townsend, 1972).

### Table 4. Research design

<table>
<thead>
<tr>
<th>Research Design</th>
<th>No. of Studies</th>
<th>% (n=83)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single group</td>
<td>46</td>
<td>55.4</td>
</tr>
<tr>
<td>Single case study</td>
<td>21</td>
<td>25.3</td>
</tr>
<tr>
<td>Two groups</td>
<td>12</td>
<td>14.5</td>
</tr>
<tr>
<td>Multiple groups</td>
<td>4</td>
<td>4.8</td>
</tr>
</tbody>
</table>

### Measurement for speech

Information about the measurement for speech is shown in Table 5. Perceptual analysis was the most common method for speech analysis and occurred in 78 (94.0%) studies. The other common method for assessing the speech ability was acoustic analysis, which occurred in 23 studies (27.7%). Imaging was used in 11 (13.3%) studies. The imaging techniques included cinefluorography, videofluoroscopy, and X-ray. Palatography occurred in six (7.2%) studies. Apart from these measures, two (2.4%) studies utilized patient filled-out
questionnaire to assess speech ability.

The variety of measurement used in a study was also investigated. In the 83 studies, thirty (36%) used more than one type of measurement to assess the speech ability. Among these 30 studies, the most common combination of measurement involves both perceptual and acoustic analysis and these were used together in 20 studies. Eleven studies involve both perceptual analysis and imaging techniques.

Table 5. Measurement used to reflect speech ability

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Included in No. of Studies</th>
<th>% (n=83)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptual analysis</td>
<td>78</td>
<td>94.0</td>
</tr>
<tr>
<td>Acoustic analysis</td>
<td>23</td>
<td>27.7</td>
</tr>
<tr>
<td>Imaging technique</td>
<td>11</td>
<td>13.3</td>
</tr>
<tr>
<td>Palatography</td>
<td>6</td>
<td>7.2</td>
</tr>
<tr>
<td>Others</td>
<td>2</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Speech samples

Table 6 shows the type of speech sample being used and the number of times it was utilized among the studies. The most common type of speech sample was single word, which were derived from word list or carried out as in an articulation test, and it was included in 29 (34.9%) studies. Sentences were used in 26 (31.3%) studies. Connected speech was used in 23 (27.7%) studies. Nonsense syllables occurred in 12 (14.5%) studies and vowels in five (6.0%) studies. Twenty-eight (33.7%) studies did not specify the speech material used for evaluating the speech.

Table 6. Types of speech samples in studies

<table>
<thead>
<tr>
<th>Speech samples</th>
<th>Included in No. of Studies</th>
<th>% (n=83)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowels</td>
<td>5</td>
<td>6.0</td>
</tr>
<tr>
<td>Nonsense syllables</td>
<td>12</td>
<td>14.5</td>
</tr>
<tr>
<td>Single word</td>
<td>29</td>
<td>34.9</td>
</tr>
<tr>
<td>Sentence</td>
<td>26</td>
<td>31.3</td>
</tr>
<tr>
<td>Connected speech</td>
<td>23</td>
<td>27.7</td>
</tr>
<tr>
<td>Unspecified</td>
<td>28</td>
<td>33.7</td>
</tr>
</tbody>
</table>

In the 83 studies, twenty-one (25.3%) used a mixture of two types of speech sample. Ten studies (12.0%) utilized three or more samples.
Perceptual parameters

A summary of the types of perceptual parameters used is shown in Table 7. Intelligibility was the most common perceptual parameter employed and was included in 59 (75.6%) studies. The other common parameters included phonetic transcription (18, 23.1%) and articulation test score (15, 19.2%). Speech rate was assessed in five (6.4%) studies. Other perceptual parameters included the Fletcher time-by-count test (Fletcher, 1972), understandability, the mean percent perceived correct production of consonant phonemes, communication effectiveness and communicative capability for conversational speech.

<table>
<thead>
<tr>
<th>Perceptual parameters</th>
<th>Included in no. of Studies</th>
<th>% (n=78)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intelligibility</td>
<td>59</td>
<td>75.6</td>
</tr>
<tr>
<td>Phonetic transcription</td>
<td>18</td>
<td>23.1</td>
</tr>
<tr>
<td>Articulation test score</td>
<td>15</td>
<td>19.2</td>
</tr>
<tr>
<td>Speech rate</td>
<td>5</td>
<td>6.4</td>
</tr>
<tr>
<td>Severity</td>
<td>4</td>
<td>5.1</td>
</tr>
<tr>
<td>Acceptability</td>
<td>2</td>
<td>2.6</td>
</tr>
<tr>
<td>Other perceptual parameters</td>
<td>7</td>
<td>9.0</td>
</tr>
</tbody>
</table>

Thirty-eight (48.7%) studies that employed perceptual measurement employed a mixture of parameters. Two (2.6%) of the studies did not specify the perceptual parameters employed but only used general impression to describe the speech.

Measurement of intelligibility

Intelligibility was the most commonly used perceptual parameter and occurred in 59 (76.6%) of the studies using perceptual analysis. In these studies, a rating scale was most common used for measuring intelligibility and occurred in 23 (39.0%) studies. Twenty of these rating scale were an equal-appearing interval scale. Twenty-two (37.3%) studies used a transcription task to measure intelligibility. Three studies (5.1%) utilized both rating scale and transcription task to measure intelligibility. A relative large number of studies (14, 23.7%) only provided a crude estimation of intelligibility. For example, the patients might be described as “intelligible”. In some cases, the intelligibility was described according to
severity (mild to severe) without explanation of the rating.

**Number and Type of listeners**

Type of listener for those studies using perceptual analysis is shown in Table 8. The majority of studies (23, 29.5%) used qualified speech-language pathologists as listeners. Twenty-one (26.9%) studies used naïve listeners. Eleven (14.1%) studies used trained listeners. A relatively large number of studies did not specify the type of listeners involved (25, 32.1%). In 7 (9.0%) studies, some of the listeners were the authors of the paper.

### Table 8. Types of listeners in studies with perceptual analysis

<table>
<thead>
<tr>
<th>Types of listeners</th>
<th>Included in no. of Studies</th>
<th>% (n=78)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech-language-pathologist (or equivalent)</td>
<td>23</td>
<td>29.5</td>
</tr>
<tr>
<td>Naive listener</td>
<td>21</td>
<td>26.9</td>
</tr>
<tr>
<td>Trained listener</td>
<td>11</td>
<td>14.1</td>
</tr>
<tr>
<td>Surgeon or Medical Staff</td>
<td>8</td>
<td>10.3</td>
</tr>
<tr>
<td>Family members</td>
<td>5</td>
<td>6.4</td>
</tr>
<tr>
<td>Graduate student SLP</td>
<td>2</td>
<td>2.6</td>
</tr>
<tr>
<td>Unspecified</td>
<td>25</td>
<td>32.1</td>
</tr>
</tbody>
</table>

The number of listeners employed to assess the speech is shown in Table 9. Of the 78 studies, only 51 (65.4%) studies specified the number of listeners involved. A relatively large number of studies (15, 19.2%) used two listeners to do so. Two studies used only one listener to evaluate speech.

### Table 9. Number of listeners

<table>
<thead>
<tr>
<th>No. of Listeners</th>
<th>No. of Studies</th>
<th>% (n=78)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>5.1</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>19.2</td>
</tr>
<tr>
<td>3-10</td>
<td>24</td>
<td>30.8</td>
</tr>
<tr>
<td>13+</td>
<td>8</td>
<td>10.3</td>
</tr>
<tr>
<td>Unspecified</td>
<td>27</td>
<td>34.6</td>
</tr>
</tbody>
</table>

In 49 (62.8%) studies, the mode of presentation of the speech sample to the listeners was unspecified. Audiotape was used in 31 (39.7%) studies and videotape was used in 6 (7.7%) studies. In two (2.6%) studies, the subjects spoke face to face to the listeners.

Blinding was defined as not having the listener aware the identity of the speakers or the condition of the speech sample. For example, blinding was carried out when the judges
were not aware whether the speech stimuli were made preoperatively or postoperatively, or when the presentation of stimuli were randomized. Among the 78 studies utilizing perceptual analysis, only 8 (10.3%) of the studies employed blinding for the listeners.

Reliability Measures

When perceptual measurement was used, only of 11 (14.1%, n=78) studies included a report of reliability. The definition of reliability used in the coding of this item included interjudge or intrajudge reliability, reliability reported by a consensus model, or reliability established previously by members of the same team. Among these 11 studies, four used a mixture of both inter-rater and intrarater reliability. Five used only interrater reliability. No studies used intra-rater reliability alone. Two studies used a consensus model (Furia et al., 2001; Leonard & Gillis, 1990).

Statistical evidence

Table 10. shows the statistical analysis utilized in the studies. Only 19 (22.9%) of the total number of studies used statistical evidence to support their conclusion. Among these studies, \( t \) test was most often used and occurred in six studies. Wilcoxon test and the Mann-Whitney test were also common statistical analysis employed and occurred in four studies respectively. One of the studies did not specify the statistical test being used for the statistical analysis and only the P-value was given.

<table>
<thead>
<tr>
<th>Statistical analysis</th>
<th>Included in No. of Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t ) test</td>
<td>6</td>
</tr>
<tr>
<td>Wilcoxon test</td>
<td>4</td>
</tr>
<tr>
<td>Mann-Whitney test</td>
<td>4</td>
</tr>
<tr>
<td>Others</td>
<td>8</td>
</tr>
</tbody>
</table>

The first focus of this critical review is on methodological considerations of speech measurement in studies of glossectomy. The second focus is a summary of the findings from these studies. This analysis follows.
Articulatory pattern after glossectomy

After glossectomy, patients were found to develop compensatory articulation strategies for their speech production. LaRiviere, Seilo, & Dimmick (1975) studied the compensatory articulation pattern in a patient with partial glossectomy. They found that subjects tended to produce speech sounds that requires tongue movement using articulation gestures associated with lips and teeth. For example, in producing the alveolar plosives /t, d/, the speakers used articulatory gestures that was similar to the production of the labial plosives /b, p/ with the lips coming to each together. In producing the alveolar fricatives /s, z/, the speakers blew the air through closed teeth instead. Georgian, Logemann, & Fisher (1982) also investigated compensatory articulation in a patient with partial glossectomy. They concluded from the videofluoroscopy, that the patient produced stop consonants by using different degree of labial protrusion and retraction depending on the vowel context.

Even though the patients might develop compensatory articulation after glossectomy, their speech in general is dependent on several factors: the extent of the surgery, the surgical procedure, the use of prosthesis, and implementation of speech therapy. Findings associated with each of these factors are summarized and analyzed in the following section.

(1) Effect of extent of surgery on speech

Among the studies being investigated, 32 (38.6%) of these investigated the effect of extent of surgery on speech production after glossectomy.

Several authors have suggested that the effect of glossectomy on speech production could be roughly estimated by the amount of tongue and adjacent oral tissue removed. For example, Massengill, Maxwell & Pickrell (1970) investigated the articulation of two patients with partial glossectomy and one patient with total glossectomy. They observed that speech become increasingly distorted as larger percentages of the tongue were removed. Rentschler & Mann (1980) studied the effects of glossectomy on intelligibility in 20 glossectomized patients. They found that the extent of surgery was negatively correlated (-.726) with overall
intelligibility \( (P<.001) \) and concluded that the effect of glossectomy on speech production depends on the extent of the surgery. Pruszewicz & Kruk-Zagajewska (1984) studied the articulation pattern in seven patients with partial glossectomy. They also observed that the deviation in speech production depended on the extent of surgical excision of the tongue.

However, other studies reported that speech is affected more by the mobility rather than the volume of the residual tongue. For example, Imai & Michi (1992) assessed articulatory function after resection of the tongue and floor of the mouth in 17 subjects using both palatometry and perceptual evaluation. They found that some patients could produce less distorted sounds even though a large part of the tongue was removed and that the lingual-palatal contact revealed by the palatometry was different from normal. On the other hand, even though some patients could form complete constriction between the tongue and the hard palate, their production of the alveolar stop \(/t/\) was still highly distorted. The authors contributed the distortion of the \(/t/\) sound to the inability of the tongue tip to release the constriction quickly. Therefore, the authors suggested that the mobility of the tongue rather than the tongue volume affects speech production after glossectomy.

(2) Effect of surgical or reconstructive procedures on speech

According to Prince & Bailey (1999), after excision of all but the smallest lesions of the anterior tongue, reconstruction is required to maintain function. The ideal reconstruction would restore the normal tongue volume and shape, have a supple non-hairy surface, and move freely in function. Reconstruction techniques can include primary closure (sewing remaining tissue together) and split-thickness skin graftings, and different types of flaps such as local, distal, or microvascular free flaps (Lazarus, 1998).

Among the studies being investigated, 31 (37.3%) studies focused on the effect of surgical procedures on speech. These studies tried to compare the effectiveness of different techniques on the speech outcome. Their main focus was on the use of the advantages different distant flaps on the functional outcome. However, even though a number of studies
had been carried out in this area, it is difficult to make a conclusion on the effect of surgical procedures on speech. It is because from the studies being reviewed, 20 (64.5%) used only a single subject design or single group design, and only nine (29.0%) used a multiple group design. Therefore, it was difficult to compare the effect of different reconstruction techniques on speech because the surgery tends to be the same for the patients. In those studies that investigate the effectiveness of a certain reconstruction method on speech after glossectomy using a single case or single group design, the researchers usually compared their result against other studies. However, as the method of speech measure was not the same among the studies, it would be difficult to come to a conclusion about the effectiveness of the method. The fact that most studies only used a single case or single group design might be due to the fact that it was difficult to recruit enough number of subjects to carry out a two group design. In this case, the cooperation of different institutions such as the study by Logemann et al. (1993) might be useful for recruiting more subjects for a two or multiple group design.

Apart from the inadequate research design, the description of the procedures in data collection was also unclear. Specifically, in 31 of the studies focusing on surgical procedures on speech, 17 (54.8%) had not specified the speech sample used. This lack of clarity in the description of the speech measures also affects the interpretation of the result.

(3) Effect of different prosthesis on speech production after glossectomy

After glossectomy, prostheses can be used to assist speech production by reshaping or reducing the volume of the oral cavity (Leonard & Gillis, 1990). There are two main types of prostheses: 1) maxillary reshaping prostheses, which lowers the palate to allow the tongue to come into contact with the palate for the production of certain speech sounds; and 2) mandibular reshaping prostheses, which provides constriction of the vocal tract so as to approximate vocal tract property in normal speech (Leonard & Gillis, 1990).

The selection of the prosthesis depends on the type and extent of surgery. For total glossectomy, the mandibular prosthesis was used. For partial glossectomy, a maxillary
prosthesis was used. Sometimes, the maxillary prosthesis may be used as a training prosthesis
and used temporarily, but in some patients, this may be used permanently (Aramany, Downs,
Beery, & Aslan, 1982).

Among the studies being investigated, 15 (18.1%) focused on the effect of prostheses
on speech. In general, the use of prosthesis leads to better speech outcome. For example,
Dholam, Bachher, & Rao (2001) studied the effect of mandibular tongue prosthesis on a
patient with total glossectomy. Speech was measured by perceptual analysis supported by
acoustic analysis. Broad transcription was carried out by a trained and experienced listener.
With the prosthesis, the patient’s intelligibility improved from 6 to 3 on a 7-point
intelligibility scale. Therefore, the authors concluded that the use of tongue prosthesis was
helpful on the subject’s speech production.

However, when concluding the effect of prosthesis on speech, the same problem
described previously was found. Among the 15 studies focusing on effect of prosthesis, 12
(80.0%) used only a single subject design and 3 (20.0%) used a single group design. No study
had used a two-group or multiple group design. When a single case or single group research
design was used, they seldom involve a proper single case study design with ABA design.

As suggested by Casper (1998), a prosthesis may affect speech production in different
ways, depending on the amount of residual mass, mobility of the tongue, extent of surgery
and other factors. Therefore, research studies need to utilize appropriate research design and
valid speech measurement so that an appropriate prosthesis could be chosen for the patient.

(4) Effect of therapy on speech production after glossectomy

Among the studies reviewed, only four (5%) focused on the effect of speech therapy
on speech after glossectomy. This finding echoed the previous conclusion of Amerman &
Laminack (1974), who also suggested that few studies focused on the effect of speech
therapy on speech after glossectomy. The authors had urged that more investigation should be
carried out to provide a base of interpretation and prediction of the treatment effect on
speech. It was surprising that after 20 years, there were still few studies on this area.

Among the few studies that reported speech therapy in patients with glossectomy, it was shown that speech therapy generally lead to better speech outcomes in the patients.

Skelly et al. (1972) studied the changes in intelligibility in ten glossectomees before and after therapy. The treatment aimed at lengthening vowel duration, reducing vocal intensity, using meaningful pauses, elevating low habitual pitch, widening pitch range, increasing the variety of pitch patterning and extending the harmonic range. Patients were trained to use compensatory articulation. The subjects were compared with ten controls. They found that the intelligibility scores for the treatment group increased more than the control group. Therefore, they concluded that their treatment was effective for the patients.

More recently, Furia et al. (2001) studied 27 patients after glossectomy and their progress after speech rehabilitation. Patients with total and subtotal glossectomy showed significant improvement in their speech intelligibility (P<.05). The authors also observed improvement in the speech intelligibility for patients with partial glossectomy, but the improvement was not significant. The authors therefore concluded that speech therapy effectively improved speech intelligibility of patients undergoing glossectomy.

The above studies provided positive examples of the beneficial effect of speech therapy after glossectomy. However, as only a few number of studies (four) had focused on the effect of speech therapy on speech after glossectomy, it is difficult to conclude whether speech therapy or what aspect of the speech therapy was beneficial to the patients.

As discussed above, rehabilitation of speech is important to the patient’s quality of life after the surgery. In order to restore functional speech in these patients, more research should be carried out on the effect of speech therapy on speech after glossectomy. When investigating the effect of speech therapy, different aspects of speech could be assessed. The mode of presentation of the patient’s speech should also be taken into consideration. For example, if compensation were used, the intelligibility would be better assessed both with
and without visual cue for comparison. Georgian, Logemann & Fisher (1982) suggested that speech therapy helps the patients develop compensatory strategies to produce speech sounds in a new way. They also suggested the patients should be taught strategies for improving communicative efficiency, such as maintaining good eye contact with listeners and speaking at a slower rate. Suggestions for speech therapy were also provided by Greven, Meijer & Tiwari (1994). They suggested that apart from developing new places of articulation, awareness of articulation could be increased through auditory feedback training and alternative communication systems could be introduced. They also suggested that modification on other aspects of speech, such as the intonation, rhythm and speech rate, could be carried out. It is recommended that more researches should be carried out on the effect of speech therapy on the speech outcome through modification of different aspects of speech.

Summary

The above were the main findings in the field regarding the speech after glossectomy. As noted, a conclusive summary could not be generated from these studies because of the limitation in their research design and their lack of clarity in the description of their data collection. Mainly, the limitation in the research design could be seen in studies on the effect of prosthesis on speech outcome because no studies concerning this factor utilized a two group or multiple group design. Description of the procedures in data collection was also unclear. For example, unclear descriptions of the speech measures were found in the studies on surgical procedures. Apart from the limitation in research design and lack of clarity in the description of procedure, the control of reliability was also neglected in most study, as shown in the result from the coding. When reliability of the measurement was not controlled, it would be difficult to generate a conclusion from the result. These issues will be further discussed in the following section.

Discussion

In order to evaluate the impact of glossectomy surgery and rehabilitation on speech, it
is essential to have rigorous studies, with adequate attention to validity, reliability and possible biases with highest “level of evidence” possible. Therefore, the main aim of this study is to investigate how the field of literature reports the speech outcomes in patients receiving glossectomy. Specifically, it was hoped that through this critical review of the literature, we could increase our understanding on how speech were measured after glossectomy so that we could define the future direction for research in this field.

One main concern raised from the review was related to the research design employed. For single case study, the treatment effectiveness could not be studied by a single pre-post test and either ABAB design, multiple baselines, or long-term follow-up after treatment are necessary at minimum to study a new treatment in individuals (Ludlow, 1991). However, even if a treatment was proved to be effective for a single subject, its effect regarding a group was still to be proved. Therefore, randomized group designs are necessary to show the effects of a treatment programme (Ludlow, 1991). In the current review, most studies utilized a single case study (21, 25.3%) or a single group design (46, 55.4%). Among the studies that involved different group of subjects, only one was randomized independent group study which assign the subjects randomly to different groups. Due to the difficulty in recruiting adequate number of subjects in a research, it might be difficult for researchers to carry out randomized group design in the study of speech outcome after glossectomy. However, it is recommended that researchers should try their best to utilize a randomized group design for their future studies. If a randomized controlled trial could not be carried out, it was suggested that an ABAB design or multiple baseline approach should be used to study the treatment effectiveness of their treatment approach.

Apart from the type of research design, the measurement for the speech ability should also be chosen with care. In this review, one of the main concerns raised related to the measurement for speech was the validity of measures used to assess speech ability. Kent (1996) suggested that perceptual judgments are subjected to a number of influencing factors,
and the conditions in which such judgments are made should be specified in reports. However, from the current review, it was found that the procedures for perceptual analysis were frequently not clearly described, which supported the previous comments of different authors (Amerman & Laminack, 1974; Pauloski et al., 1993; Prince & Bailey, 1999). For example, about 34% of studies did not specify the speech material used for evaluating speech. When intelligibility was used for speech analysis, a relative large number of studies (24%) did not specify the method of measuring intelligibility. In addition, two (3%) studies only used a general comment to describe the speech, with no specific observations other than saying that the patient could quickly regain the speech ability. This crude estimation was clearly not a valid measurement for speech. As suggested above, the unclear description of the procedure in measuring speech ability affects the interpretation of the results. It would also be difficult to compare results across studies because the procedures used were unclear.

In order to make the speech measurement valid, bias should be controlled. From the review, it was found that there was a lack of control of bias in many of the studies. When perceptual analysis was used, only 8 (10.3%) of the studies employed blinding for the listeners. During perceptual analysis, the judgment of the listener might be affected if the identity of the speaker was known (Kent, 1996). Therefore, blinding was an important procedure to employ in perceptual study so that the listener would not be affected by the identity of the speaker. Blinding could be carried out by randomizing the order of occurrence of the speech sample or by not letting the listener know the condition of the speech sample. Wyatt et al. (1996) suggested that speech therapists not involved with the treatment of the subject should be used as raters when studying the speech outcomes in cleft palate. The same applies to the study of speech after glossectomy and the authors should not be the listeners for the speech assessment. However, in this review, seven studies (9.0%) had the authors acting as the raters for the assessment of speech ability. Listener bias will definitely occur in this case because the researcher might tend to support or was against a certain kind of approach.
For a valid measure of speech, the tool for measuring the speech should be chosen so that it is valid and truly reflect the speech ability of the patients. When reviewing research studies on assessment of intelligibility in cleft palate speech, Whitehill (2002) found that interval scaling was used in a large number of studies although it has been shown that this might not be a valid measurement of speech intelligibility. In this review, intelligibility was used as the most common perceptual assessment of speech ability after glossectomy and occurred in 75.6% of the studies with perceptual analysis. Among these studies, twenty studies employed interval scaling for assessing intelligibility. It is recommended that attention should be paid to the selection of appropriate tools for assessment of speech ability. For example, it was suggested by Whitehill (2002) that the use of phonetic transcription or a forced choice method was more valid than an equal appearing interval scale.

Apart from the validity issue, reliability was also a main concern raised in the review. Reliability reported the agreement of the ratings in repeated measure (Shaughnessy, & Zechmeister, 1997) and it could be established through inter-rater reliability or intra-rater reliability. For studies which utilized only one judge, intra-rater reliability should be reported. To establish the inter-rater reliability, more than one rater should be involved. In the current review, four (5.1%) involved one listener only and among these studies, intra-rater reliability was not reported. On the other hand, 60.3% of the studies included more than one listener. However, 80.9% of these studies did not report inter-rater reliability. This finding was similar to those found by Whitehill (2002) and Lohmander & Olsson (2004) in their review of measurement in cleft palate speech. As perceptual assessment was affected by a number of variables, including the severity of the speech disorder, the speech stimuli used, listeners’ experience with the disorder and the speech sample, and the measurement task employed (Whitehill, 2002), the researchers should try to control these variables and investigate the reliability of their measures before they claimed that a certain kind of treatment method was effective.
When assessing the speech ability by perceptual analysis, the mode of presentation should also be considered. Bosman et al. (1991) found that the speech of individuals with oral cancer was understood better if presented under video rather than audio conditions. In the current review, 49 (62.8%) studies did not specify the mode of presentation of the speech sample to the listeners. As suggested by Lohmander & Olsson (2004), even though it is controversial whether the mode of presentation affects the speech measures, tape recordings (audio or video) should be used in research for better reliability and for future reference.

In this review, two recent studies (Bressmann, Whitehill, Sader, Samman, & Hoole, 2002; Bressmann, Whitehill, Sader, & Samman, in press) used surgical mapping to describe the surgical procedures. Beck et al. (1998) had suggested that the literature studying the speech outcome after glossectomy usually described their surgical pattern in general terms and the unclear description prevent the understanding of the relationship between the surgical patterns with the speech outcome. Therefore, they suggested the use of mapping to reflect more precisely the surgical resection carried out. They developed their mapping protocol from clinical trials with over 100 patients. They used external views of the face to indicate the route of surgical entry and intra-oral views to indicate areas of tissue excised or damaged within the oral and pharyngeal cavities. Using this protocol, they provided a clear description of the site and volume of resection of the tongue. It is recommended that future studies could use this method for documenting more precisely the extent and location of surgery to allow detail analysis of the relationship between the speech outcome and the surgery.

In this review, it was also found that perceptual analysis was the most common (78, 94.0%) measurement for speech. Perceptual analysis was an appropriate measurement of speech ability. However, to increase our understanding about the speech production, researchers might like to use instrumental analysis to supplement the perceptual analysis (Kent, 1996). In this way, the compensatory articulation pattern of the client might be documented and speech therapy might be carried out with these instruments. In the current
review, instruments had been used in documenting the speech of patients with glossectomy. For example, as reported by Fletcher (1988), palatography could be used to identify the different articulatory pattern of speakers with glossectomy. However, this method might also be utilized for speech therapy. In the current review, there were four studies on speech therapy after glossectomy and these studies had utilized perceptual and acoustic analysis to document the speech production. None of these studies utilized palatography for documenting the effect of speech therapy. This technique had been utilized in the treatment of other speech patients. For example, Hardcastle, Gibbon, & Jones (1991) had provided detailed description on how researchers could use this technique to carry out treatment for patients with cleft palate or phonological disorder. To increase our understanding on speech therapy on glossectomy speech, research on the speech therapy using palatography might be a possible direction for future research.

Statistical analysis was used in 22.9% of the studies to support their conclusion. When researchers were assessing the effect of a particular treatment technique on the speech after glossectomy, it is suggested that statistical analysis might be employed so that further statistical evidence could be obtained.

In this review, it was found that there was a lack of unity in the procedures and materials for assessing speech ability after glossectomy. In order to compare results across studies, the speech materials and methods used should be similar. Therefore, it is recommended that researchers try to develop a standard protocol for assessing speech after glossectomy. With a similar assessment protocol, researchers could compare the difference in the speech outcomes of different surgical techniques or rehabilitation programmes.

Acknowledgement

I wish to express appreciation to Dr. T. Whitehill, my supervisor, for her assistance and advice on this study, and for the help with the repeated review.
References


Phonetics, 17, 411-420.


Appendix

The followings are studies meeting the inclusion criteria in for the current review.


Haughey, B.H., Taylor, S.M., & Fuller, D. (2002). Fasciocutaneous flap reconstruction of the tongue and floor of mouth - Outcomes and techniques. *Archives of Otolaryngology -


Morrish, L. (1984). Compensatory vowel articulation of the glossectomee: acoustic and


