Anatomical and Functional Correlates of Voice Quality in Tracheoesophageal Speech

*†Corina J. van As-Brooks, *†‡Frans J.M. Hilgers, †Florien J. Koopmans-van Beinum, and †Louis C.W. Pols

Amsterdam, The Netherlands

Summary: The purpose of the current study was to assess the anatomic and functional correlates of voice quality in tracheoesophageal speech, with dynamic imaging studies of the neoglottis. Videofluoroscopy (providing a lateral view), digital high-speed endoscopy (providing a "birds-eye" view), and their relationships with perceptual evaluations of voice quality were investigated. Several significant relationships were found. Imaging with videofluoroscopy revealed that the following anatomic and functional parameters (established during phonation) are related to voice quality: *presence of a neoglottic bar*, *regurgitation of barium*, *tonicity of the neoglottis*, and *minimal neoglottic distance*. Furthermore, the *index of the increase of the maximal subneoglottic distance from rest to phonation* also showed a significant relationship with voice quality. Imaging with digital high-speed endoscopy revealed features relevant to voice quality, including *amount of saliva*, *visibility of the origin of the neoglottis*, *shape of the neoglottis*, and *regularity of the vibration*. Knowledge of the anatomic and functional correlates of tracheoesophageal voice quality provides prerequisite information for future (phono-) surgical and/or clinical improvements to the voice quality of postlaryngectomy (prosthetic) voice production.

Key Words: Laryngectomy—Tracheoesophageal speech—Voice quality—Perceptual evaluation—Videofluoroscopy—High-speed digital endoscopy.

INTRODUCTION

Videofluoroscopy (VF) is a useful tool for investigators researching the anatomic and functional characteristics of the neoglottis and is used by physicians widely.[1,2] During voicing especially, a good (lateral) image of the dynamics of the new sound source in persons having had laryngectomies can be obtained. After the introduction of voice prostheses in 1980,[3] several VF studies were performed to investigate possible differences between the neoglottis in esophageal and in tracheoesophageal (TE) speech and to investigate the role of the neoglottis in TE speech.[4–6] Of these 3 studies, only Van Weissenbruch et al[5]

Accepted for publication July 21, 2004.

From the *Department of Head and Neck Surgery and Oncology, The Netherlands Cancer Institute, Amsterdam, The Netherlands; †Institute of Phonetic Sciences, Amsterdam Center for Language and Communication (ACLC), University of Amsterdam, The Netherlands; ‡Department of Otolaryngology, Academic Medical Center, University of Amsterdam, The Netherlands.

Address correspondence and reprint requests to Frans J. M. Hilgers, Plesmanlaan 121, 1066CX Amsterdam, The Netherlands.

Journal of Voice, Vol. 19, No. 3, pp. 360–372
0892-1997/$30.00

doi:10.1016/j.jvoice.2004.07.011

investigated the relationship between the neoglottis and the quality of TE speech (judgment of phonatory skills, additional factors, and general judgment resulting in a combined judgment as good, moderate, or poor). They found that spasm, hypertonicity, and hypopharyngeal stricture were associated with both esophageal and TE speech failures. They also established that laryngectomized patients who underwent primary myotomy and neurectomy were most successful in obtaining good TE speech, whereas those patients who did not undergo an additional surgical procedure showed the most speech failures.

Digital high-speed (HS) endoscopy is at present only used by physicians as a research tool, although developments for clinical implementation are ongoing and promising. With HS endoscopy, which in contrast to stroboscopy is independent of the fundamental frequency of the sound source, it is possible to obtain images of the vibrating neoglottis from above (birds-eye view). The recordings have been shown to give valuable information in addition to the information obtained by videofluoroscopy.[7] van As et al[8] has described the use of HS endoscopy by physicians as a tool to study the vibratory behavior of the neoglottis. Results of that study showed large variability in several anatomic and functional characteristics of the neoglottis, such as the shape of the neoglottis, and the regularity and localization of the vibration. Before the work of van As et al,[8] no studies were performed with HS imaging to study characteristics of the neoglottis. Therefore, the anatomic and functional characteristics of the neoglottis during speech with HS imaging and their relationships with voice quality of TE speech remain to be studied. Lundström and Hammarberg[9] studied one aspect of intelligibility (the voiced–voiceless distinction) in one esophageal and one TE speaker. They concluded that the HS recordings together with the simultaneously recorded voice signal give answers to important questions about the function of the pharyngoesophageal (PE) segment (a synonym for neoglottis).

With HS endoscopy as new and still developing, the need for investigation of possible relationships between observations made with this new imaging method and voice quality is evident. Therefore, in the current study, we investigated the relationships between anatomic and functional neoglottic characteristics, as visualized with VF and HS endoscopy, and the quality of the TE voice.

PATIENTS AND METHODS

Patients

The patient group consisted of 46 patients.[8] However, for various reasons, some patients participated only in one or some parts of the study. For investigation of the relationships of the neoglottic characteristics with TE voice quality, data of 38 patients were available to study the relationships between VF and TE voice quality, and data of 38 patients were available to study the relationships between HS endoscopy and TE voice quality.

All patients used TE speech by means of an indwelling voice prosthesis.[10] They were selected from a group of 173 patients that were in follow-up at The Netherlands Cancer Institute. To obtain a selection that encounters all varieties present in the patient population, female patients and patients after pharyngeal reconstruction also were included. An overview of the patient characteristics is given in Table 1.

Methods

The auditory perceptual ratings we used in the current study consist of an overall judgment of voice quality (*good, reasonable, poor*) and more specific judgments of 17 bipolar semantic scales by 4 trained speech-language pathologists (SLPs). The scales are judged on a 7-point scale, in which 1 represents the negative and 7 the positive scale end. The auditory perceptual scales and the methods we used have been previously described.[11] The scales selected for investigating the relationships with neoglottic characteristics have been shown to have sufficient interrater reliability and agreement when judged by trained expert raters. The selected semantic scales are as follows: *deviant–normal**, *unpleasant–pleasant**, *ugly–beautiful**, *noise–no noise**, *monotonous–melodious**, *expressionless–expressive**, *weak–powerful***, *unsteady–steady**, *jerking–fluent**, *slow–quick*##, *low–high*#, *deep–shrill*#, *bubbly–not bubbly***, *breathy–not breathy***, *hypertonic–not hypertonic***, *hypotonic–not hypotonic***, and *unintelligible–intelligible**. The first three scales are general and