Acoustic Signal Typing for Evaluation of Voice Quality in Tracheoesophageal Speech

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

Amsterdam, The Netherlands

Summary: Because of the aperiodicity of many tracheoesophageal voices, acoustic analysis of the tracheoesophageal voice is less straightforward than that of the normal voice. This study presents the development and testing of an acoustic signal typing system based on visual inspection of a narrow-band spectrogram that can be used by researchers for classification of voice quality in tracheoesophageal speech. In addition to this classification system, a selection of acoustic measures [*median fundamental frequency, standard deviation of fundamental frequency, jitter, percentage of voiced (%Voiced), harmonics-to-noise ratio (HNR), glottal-to-noise excitation (GNE) ratio,* and *band energy difference (BED)*] was computed to provide more insight into the acoustic components of tracheoesophageal voice quality. For clinical relevance, relationships between the acoustic signal types and an overall judgment of the voice were investigated as well. Results showed that the four acoustic signal types form a good basis for performing more acoustic analyses and give a good impression of the overall quality of the voice.

Key Words: Acoustic analysis—Laryngectomy—Tracheoesophageal speech—Voice prosthesis.

INTRODUCTION

Voice quality is a perceptual phenomenon, and consequently, perceptual evaluations are considered the ‘‘gold standard’’ of voice quality evaluation. Disadvantages of perceptual evaluations are that listeners differ in their opinion about voice quality and that it is time consuming to acquire these judgments, because many raters are needed to obtain sufficient inter- and intrarater reliability.[1]

In clinical practice, perceptual evaluations play a prominent role in therapy evaluation purposes. Acoustic analyses are usually not routinely performed for clinical purposes. Acoustic measures do not show a one-to-one relationship with perceptual evaluation

Accepted for publication April 8, 2005.

Supported in part by a grant from the Amsterdam Center for Language and Communication, University of Amsterdam, the Netherlands; and from the Department of Speech and Language of the Radboud University Nijmegen, the Netherlands. The Maurits and Anna de Kock Foundation provided financial support for the equipment needed for the speech recordings, acoustic analyses, and the listening experiment.

From the *Department of Otolaryngology-Head & Neck Surgery, Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands, the †Institute of Phonetic Sciences, Amsterdam Center for Language and Communication, University of Amsterdam, Amsterdam, The Netherlands, and ‡Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

Address correspondence and reprint requests to Frans J. M. Hilgers, Department of Otolaryngology-Head & Neck Surgery, The Netherlands Cancer Institute/Antoni van Leeuwenhoek Hospital, Plesmanlaan 121, 1066CX Amsterdam, The Netherlands. E-mail: f.hilgers@nki.nl

Journal of Voice, Vol. 20, No. 3, pp. 355–368
0892-1997/$32.00

doi:10.1016/j.jvoice.2005.04.008

and therefore cannot serve as a substitute but merely as an adjunct to it. They can provide more insight in the acoustic characteristics of a voice. Several studies have described acoustic analyses of tracheoesophageal voice quality and have concluded that tracheoesophageal voices differ considerably from normal voices with respect to the acoustic measures.[2–5]

Acoustic measures computed digitally will always produce the same result for the same input, and when these measures are obtained correctly, they can form a valuable objective adjunct to perceptual evaluations in clinical practice. Unfortunately, acoustic analysis of the tracheoesophageal voice is less straightforward than that of the normal voice. It is because in many voices, aperiodicity is evidenced, and in some voices, the fundamental frequency is extremely low. Also, the acoustic measures must be perceptually relevant, which yet has to be shown.

In an earlier study of tracheoesophageal voice quality, moderate-to-strong correlations were found between the perceptual evaluations of a sustained /a/ and the acoustic measures, which are calculated on the same sustained /a/ with the *Multi Dimensional Voice Program* (*MDVP*) (Kay Elemetrics, Lincoln Park, NJ).[5] This study also showed that with *MDVP*, 30% of the voice samples could not be analyzed at all, or only very short parts were analyzable. Visual inspection of these voice samples showed that the patients had very low-pitched voices (and therefore fell outside the fixed pitch analysis range of *MDVP*) or had very aperiodic voices. Tracheoesophageal voices thus can be aperiodic to such an extent or can have such an extremely low pitch that the pitch detection algorithm fails, or even that there is no fundamental frequency present at all. It implies that acoustic measures based on pitch detection algorithms only provide reliable results for the tracheoesophageal voices with more regular periodicity. Narrow-band spectrograms to determine the overall acoustic character of the tracheoesophageal voice to be analyzed provide a good impression of the harmonic characteristics of the voice and consequently the ability to perform reliable periodicity-based acoustic measures.

Although for normal laryngeal voices, acoustic signal types (based on narrow-band spectrograms) are advised to be used by researchers as a visual information tool and as a decision-making tool for further acoustic analyses,[6,7] such a system has not been developed and presented for tracheoesophageal speech.

The aim of this study is to develop an acoustic signal typing system that is perceptually relevant, that can evaluate the entire range of tracheoesophageal voices, and that can serve as an underlying basis for further acoustic analyses. Acoustic measures are selected that enable calculations for the entire range of tracheoesophageal voice qualities, and subsequently, relationships between the acoustic signal types and the acoustic measures will be investigated to gain insight into the specific acoustic characteristics of the signal types. Furthermore, relationships between an overall perceptual judgment of voice quality and the acoustic signal types will be investigated. We will use these relationships to gain insight into the perceptual relevance of the acoustic signal types and to investigate whether they might form a valuable adjunct to perceptual evaluation in everyday clinical practice.

PATIENTS AND METHODS

Patients

Speech recordings were made of a total of 40 laryngectomized patients with tracheoesophageal speech by means of an indwelling voice prosthesis.[8] One speaker refused to produce a sustained /a/, which left 39 laryngectomized persons for the analyses. Twenty-nine of them were men, and 10 were women. Patient ages ranged from 47 to 82 years, with a mean of 66 years. Postoperative follow-up ranged from 1 to 18 years, with a mean of 6 years. More information about the speakers participating in this study is summarized in Table 1.

Speech material, recording, and processing

The speech material for the acoustic analyses consisted of three sustained vowels /a/ at a comfortable pitch and loudness level and a standard read-aloud text. The speech recordings were made in a quiet, sound-treated room. For the recordings, we used a DAT-recorder (Sony TCD-8; Sony Corporation, Tokyo, Japan), together with the hardware and software of the Computerized Speech Lab, Model 4300B (Kay Elemetrics, Lincoln Park, NJ). Via the external module of the Computerized Speech Lab, the speech data were digitally recorded on DAT