Towards Objective Voice Assessment: The Diplophonia Diagram

*,†Philipp Aichinger, *Imme Roesner, *Berit Schneider-Stickler, *Matthias Leonhard, *Doris-Maria Denk-Linnert, *Wolfgang Bigenzahn, †Anna Katharina Fuchs, †Martin Hagmüller, and †Gernot Kubin, *Vienna and †Graz, Austria

Summary: Objectives. Diplophonia is an often misinterpreted symptom of disordered voice, and needs objectification. An audio signal processing algorithm for the detection of diplophonia is proposed. Diplophonia is produced by two distinct oscillators, which yield a profound physiological interpretation. The algorithm's performance is compared with the clinical standard parameter degree of subharmonics (DSH).

Study Design. This is a prospective study.

Methods. A total of 50 dysphonic subjects with (28 with diplophonia and 22 without diplophonia) and 30 subjects with euphonia were included in the study. From each subject, up to five sustained phonations were recorded during rigid telescopic high-speed video laryngoscopy. A total of 185 phonations were split up into 285 analysis segments of homogeneous voice qualities. In accordance to the clinical group allocation, the considered segmental voice qualities were (1) diplophonic, (2) dysphonic without diplophonia, and (3) euphonic. The Diplophonia Diagram is a scatter plot that relates the one-oscillator synthesis quality ($SQ1$) to the two-oscillator synthesis quality ($SQ2$). Multinomial logistic regression is used to distinguish between diplophonic and nondiplophonic segments.

Results. Diplophonic segments can be well distinguished from nondiplophonic segments in the Diplophonia Diagram because two-oscillator synthesis is more appropriate for imitating diplophonic signals than one-oscillator synthesis. The detection of diplophonia using the Diplophonia Diagram clearly outperforms the DSH by means of positive likelihood ratios (56.8 versus 3.6).

Conclusions. The diagnostic accuracy of the newly proposed method for detecting diplophonia is superior to the DSH approach, which should be taken into account for future clinical and scientific work.

Key Words: Diplophonia–Dysphonia–Acoustic analysis–Signal processing–Detection.

INTRODUCTION

Diplophonia is a common and often misinterpreted symptom of disordered voice. From a clinical point of view, it is characterized by the presence of two simultaneous pitches in the voice sound.[1] Depending on the underlying etiology, the presence of diplophonia indicates the necessity for phonosurgery and/or speech therapy. The absence of diplophonia is used as an outcome measure in clinical studies for evaluating treatment techniques.[2–13] Detecting diplophonia is therefore important for treatment indication and treatment effect measurement. To cope with principles of evidence-based medicine, objective voice assessment methods are needed.[14] Thus, the presence of diplophonia should be determined objectively; but in current clinical practice, it is determined auditively by clinical experts.

Two added periodic waveforms in the audio signal may provoke the sensation of two simultaneous pitches, which is described by the basic pitch perception model.[15] This model motivates the decomposition of disordered voice into periodic waveforms and a noise component for analysis. Several attempts for decomposing voice signals into independent periodic and noise components have been proposed, but the existing procedures have not yet been trained and tested on perceptually rated clinical data.[16–24]

Ambiguous definitions are a problem in classifying nonmodal phonation[25,26] and hinder the scientific and clinical communication. In contrast to the perceptual definition of diplophonia, there is also a definition regarding the shape of the waveform.[27] To obtain a unique definition of diplophonia, an algorithm for the automatic detection of diplophonic episodes from audio recordings taken during rigid telescopic high-speed video laryngoscopy is proposed. Defining detection methods for different kinds of nonmodal phonation will help in understanding and unifying the terminology of observed voice phenomena.[28–35]

Because subharmonics[27] are often found in diplophonic voice, the clinical standard parameter degree of subharmonics (DSH)[36] is compared with our new algorithm. Subharmonic frequencies are defined as "frequencies that lie between or below the harmonic frequencies and are rational division of the fundamental frequency or their integer multiple."[27] The DSH aims at detecting subharmonic frequencies, and its relation to perceptively determined diplophonia is investigated in the present study.

The aim of the study was to propose, test, and evaluate an analysis-by-synthesis audio signal processing algorithm for the objective detection of diplophonia from sustained phonation during rigid telescopic high-speed video laryngoscopy.

SUBJECTS AND METHODS

Clinical data collection and group allocation

A total of 50 subjects with dysphonia (28 with diplophonia and 22 without diplophonia) and 30 subjects with euphonia have been included in the study. The subjects with dysphonia were

Accepted for publication November 25, 2014.

Parts of this study were presented at the 2014 Annual Symposium of The Voice Foundation: Care of the Professional Voice.

From the *Department of Otorhinolaryngology, Division of Phoniatrics-Logopedics, Vienna, Medical University of Vienna; and the †Signal Processing and Speech Communication Lab, Graz University of Technology, Graz, Austria.

Address correspondence and reprint requests to Philipp Aichinger, Department of Otorhinolaryngology, Division of Phoniatrics-Logopedics, Medical University of Vienna, Waehringer Guertel 18–20, 1090 Vienna, Austria. E-mail: philipp.aichinger@meduniwien.ac.at

Journal of Voice, Vol. 31, No. 2, pp. 253.e17–253.e26
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.06.021

TABLE 1.
Clinical Group Allocation Versus Diagnosis of Dysphonic Subjects

		Clinical Group		
		Diplophonic	Dysphonic	Sum
Diagnosis	Laryngitis (acute/chronic)	2	8	10
	Sulcus glottidis	2	2	4
	Nodules	1	1	2
	Polyp	0	1	1
	Edema	5	3	8
	Cyst	2	2	4
	Scar	1	1	2
	Paresis	9	1	10
	Functional dysphonia	4	3	7
	Benign tumor	1	0	1
	Bamboo nodes	1	0	1
	Sum	28	22	50

recruited from the outpatients of the Medical University of Vienna, Austria, Department of Otorhinolaryngology, Division of Phoniatrics-Logopedics. The clinical group allocation has been performed by medical doctors during clinical examination, based on perceptual impression of connected speech. The subjects with euphonia were recruited via public announcement in Vienna, Austria. From each subject, several sustained phonations have been recorded during rigid telescopic high-speed video laryngoscopy, resulting in a total number of 185 phonations (approximately 380 s). An AKG HC577L (AKG Acoustics GmbH, Vienna, Austria) omnidirectional head-worn condenser microphone and a TASCAM DR-100 (TEAC CORPORATION, Tokyo, Japan) handheld audio recorder were used. The sampling rate was 48 kHz, and the quantization resolution was 24 bits. The data collection has been approved by the institutional review board of the Medical University of Vienna, Austria.

Table 1 shows the diagnoses of the subjects with dysphonia with respect to clinical group allocation.

Segmental voice quality labeling and segment selection

We distinguish between two ways of determining the presence of diplophonia. The first way is auditive rating by medical doctors, which provides one label for each subject. The second way is the labeling regarding phonation segments, which considers temporal changes of the presence of diplophonia within sustained phonation. All recorded phonations were perceptively labeled and split up into segments of homogeneous voice quality by the first author (PA). The labeling was performed on the audio material played back through AKG K702 (AKG Acoustics GmbH, Vienna, Austria) headphones. Only the audio waveforms and spectrograms were visible to the rater. In accordance to the clinical group allocation, the considered voice quality labels were (1) euphonic, (2) diplophonic, and (3) dysphonic. Figure 1 shows an example of recorded phonation, labeled for segmental voice quality. The phonation is dysphonic at first, diplophonic in the middle, and dysphonic again later. Thus, the shown example was split up into three segments for separate analysis. The

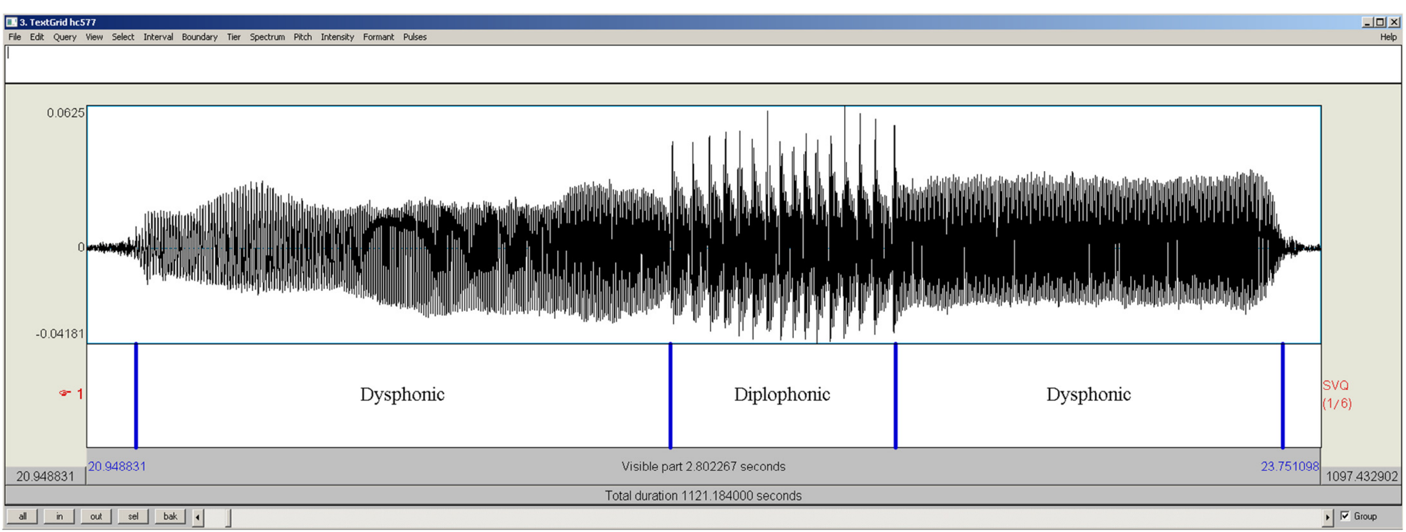

FIGURE 1. Example of a recorded phonation, labeled for segmental voice quality. The phonation is dysphonic at first, diplophonic in the middle, and dysphonic again later.