Maximum Phonation Time: Variability and Reliability

***Renée Speyer, †Hans C. A. Bogaardt, ‡Valéria Lima Passos, §Nel P. H. D. Roodenburg, *Anne Zumach, *Mariëlle A. M. Heijnen, *Laura W. J. Baijens, *Stijn J. H. M. Fleskens, and *Jan W. Brunings,** **Amsterdam and *‡§Maastricht, †Amsterdam, The Netherlands*

Summary: The objective of the study was to determine maximum phonation time reliability as a function of the number of trials, days, and raters in dysphonic and control subjects. Two groups of adult subjects participated in this reliability study: a group of outpatients with functional or organic dysphonia versus a group of healthy control subjects matched by age and gender. Over a period of maximally 6 weeks, three video recordings were made of five subjects' maximum phonation time trials. A panel of five experts were responsible for all measurements, including a repeated measurement of the subjects' first recordings. Patients showed significantly shorter maximum phonation times compared with healthy controls (on average, 6.6 seconds shorter). The averaged interclass correlation coefficient (ICC) over all raters per trial for the first day was 0.998. The averaged reliability coefficient per rater and per trial for repeated measurements of the first day's data was 0.997, indicating high intrarater reliability. The mean reliability coefficient per day for one trial was 0.939. When using five trials, the reliability increased to 0.987. The reliability over five trials for a single day was 0.836; for 2 days, 0.911; and for 3 days, 0.935. To conclude, the maximum phonation time has proven to be a highly reliable measure in voice assessment. A single rater is sufficient to provide highly reliable measurements.

Key Words: Voice–Maximum phonation time–Reliability–Interrater reliability–Intrarater reliability–Repeated measures.

INTRODUCTION

Voice is a multidimensional phenomenon.[1] The multidimensional voice assessment proposed by the European Laryngological Society includes aerodynamic measurements.[2] Maximum phonation time is usually used for practical reasons: determination of maximum phonation time is a noninvasive, fast, and low-budget measurement. The ability to maximally sustain a vowel after having taken a maximal inspiration is considered an objective measure of the efficiency of the respiratory mechanism during phonation.[3] In the literature, maximum phonation time has been used to objectify the degree of severity of dysphonia and to determine the effects of voice therapy.[4] However, some authors question the usefulness of this measure as an evaluation tool in therapy, as no significant change could be demonstrated after therapy in a group of patients with vocal nodules (eg, Treole and Trudeau[5]).

Many studies have provided normative data on maximum phonation time in diverse subject populations (see for an overview, Baken and Orlikoff[6]). The variability is high, depending on subject characteristics, such as gender or age[7] and testing conditions.[3,8] In 1968, Hirano et al stated that although tests, such as maximum phonation time, indicate the degree of vocal function, no diagnosis of laryngeal disease could be established,[9] an assertion that has since been confirmed by other authors.[10,11] In a study on procedural aspects of eliciting maximum phonation time, Neiman and Edeson[3] concluded that subjects should receive a complete verbal and visual model of the experimental task before its elicitation; instructions should be standardized; and at least three trials of maximum phonation time were required before performance could be expected to approximate the criterion. However, in a study by Lewis et al,[7] the authors found that using three trials to determine maximum phonation time in children yielded inadequate data, as most subjects had not yet reached their maximum achievement after three trials. Using more trials, higher achievement levels were found. On the other hand, Finnegan[12] demonstrated the presence of fatigue effects after practice effects. Furthermore, Shanks and Mast[13] considered the differential operation of fatigue to be partially supported by the progressive increase in standard deviations found when raising the number of maximum phonation trials. Thus, the outcomes of studies on fatigue versus practice effects while performing maximum phonation tasks are not quite consistent. Although quite a few studies have described maximum phonation time in diverse subject populations and under various testing conditions, limited information is available on the reliability of the data over time.[7]

To our knowledge, no study has thus far determined how many days the subjects should be repeatedly measured, nor have any authors indicated how many trials and raters would be necessary to obtain reliable maximum phonation time measurements using dysphonic subjects and control subjects matched by gender and age. The purpose of the present study is to determine the reliability of maximum phonation time as a function of the number of trials, days, and raters in dysphonic and control subjects.

Accepted for publication October 2, 2008.

From the *Department of Otorhinolaryngology and Head and Neck Surgery, Maastricht University Medical Centre, Maastricht, The Netherlands; †Department of Otorhinolaryngology, Academic Medical Centre, Amsterdam, The Netherlands; ‡Department of Methodology and Statistics, University Maastricht, Maastricht, The Netherlands; and the §Department of Neurology, Maastricht University Medical Centre, Maastricht, The Netherlands.

Address correspondence and reprint requests to Renée Speyer, Department of Otorhinolaryngology and Head and Neck Surgery, Maastricht University Medical Centre, P.O. Box 5800, 6202 AZ, Maastricht, The Netherlands. E-mail: renee.speyer@mumc.nl

Journal of Voice, Vol. 24, No. 3, pp. 281-284
0892-1997/$36.00

doi:10.1016/j.jvoice.2008.10.004

METHODS

Subjects

This study was conducted on two groups of adult subjects: patients with functional or organic dysphonia versus healthy subjects who did not suffer from any voice problems. The patients were diagnosed by a laryngologist at the Otorhinolaryngology Department of the University Hospital Maastricht. The same exclusion criteria were applied to both groups: pneumopathy, including chronic obstructive pulmonary disease (COPD) or atopic syndrome; upper respiratory infection; inhalation of corticosteroids; extensive laryngeal surgery; substitution voice after laryngectomy; neuromuscular diseases; or pregnancy. All subjects had to be older than 18 years to rule out bias by voice maturation or mutation, but had to be younger than 65 years. The patients had received no voice therapy for at least 3 months before inclusion.

In total, 27 adult outpatients were included. Table 1 presents the frequency of the etiologic categories as diagnosed by a laryngologist using laryngostroboscopy. The group of patients consisted of 13 men and 14 women ranging in age from 18 to 63 years. The average age for the female participants was 32 years and for the male subjects, 52 years. A control group was matched by age and gender.

TABLE 1.
Distribution of Patients by Diagnostic Categories

Phoniatric Diagnosis	N = 27
Muscle tension dysphonia	9
Submucosal swelling	4
Vocal fold nodules	2
Vocal fold polyps	2
Unilateral vocal fold paralysis	1
Other: slight vocal fold abnormalities	7
Other: severe vocal fold abnormalities	2

Procedure

Over a period of maximally 6 weeks, three digitized video recordings were made of individual subjects' performances (Mini-DV Camera-Recorder AG-DVC30; Panasonic, Matsushita, Electric Industrial Co., Osaka, Japan). During each recording, the subjects were asked to produce a sustained vowel /a:/ for as long as possible. The subjects were allowed five trials in a row with a 15-second break between each one. Before each recording, the subjects received verbal instructions according to a strict protocol in addition to a visible and audible trial performance by one and the same researcher.

Panel of expert listeners

The panel of expert listeners consisted of one laryngologist and four speech therapists. Each panel member received a complete set of all digitized video recordings in randomized order. The maximum phonation time for all patients per trial was determined using a stopwatch. By means of computers, the experts were allowed to listen to the stimuli as often as convenient in individual listening sessions. Three to 6 weeks after doing the initial rating, all experts received the randomized first recordings once more for a repeated rating.

Statistical analysis

Variance components analysis was performed for each group separately to determine the proportions of total maximum phonation time variability that can be attributed to its different sources, that is, trial, days, subjects, and raters. To estimate the variance components, a linear mixed model (*proc mixed*; SAS Institute, Inc., version 9.1) was fit, assuming a hierarchical structure of the data, whereby trials were nested within days and days were nested within patients (a three-level model, under the premise of no interaction among the explanatory variables). Interclass correlation coefficients (ICC) and Cronbach's alpha were used as measures of agreement (absolute) and internal consistency (SPSS Inc., Chicago, USA, version 11.5). Group differences between patients versus healthy controls were determined using a paired-samples *t* test. The Spearman-Brown formula was applied to compute the minimum necessary number of trials, days, and raters to obtain reliable measurements.[14]

RESULTS

Variance components and reliability coefficients

The estimated variances of raters, subjects, days, and positions (trials) are displayed together with their *P* values in Table 2 for patients and controls, respectively. (The null hypothesis assumes the variance components to be zero.) Note that the variance attributed to "rater" did not significantly contribute to the total variances for maximum phonation time. In both groups, the subjects themselves are the predominant source of variation, accounting for approximately 80% of the total variance in maximum phonation time. The second significant source is attributed to "day" (11% and 14%), followed by "trials" (6.5% and 7.6%). Very little (about 2%) of the total variance remains

TABLE 2.
Variance Components Table for Patients and Controls

Source	Estimated σ^2	SE	*P*	% Total
Patients				
Rater	0.07492	0.05416	0.0833	0.1333
Subject	45.4401	13.2675	0.0003	80.877
Day	6.3272	1.3677	<0.0001	11.2615
Position	3.6766	0.3022	<0.0001	6.5438
Residual	0.6651	0.02362	<0.0001	1.837
Total	56.1839			
Controls				
Rater	0.06671	0.04750	0.0801	0.0849
Subject	61.2618	18.1398	0.0004	78.0182
Day	11.0112	2.3589	<0.0001	14.0230
Position	6.0011	0.4818	<0.0001	7.6425
Residual	0.1816	0.006483	<0.0001	2.1308
Total	78.5224			

Abbreviation: SE, standard error.