Variation of voice quality features and aspects of voice training in males and females

Sulter, Arend Marten

Variation of voice quality features and aspects of voice training in males and females

RIJKSUNIVERSITEIT GRONINGEN

Variation of voice quality features and aspects of voice training in males and females

Proefschrift

ter verkrijging van het doctoraat in de
Medische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus Dr F. van der Woude
in het openbaar te verdedigen op
woensdag 23 oktober 1996
des namiddags te 4.15 uur

door

Arend Marten Sulter

geboren op 25 augustus 1962
te Leeuwarden

Promotores: Prof. dr. F.W.J. Albers
Prof. dr. ir. H.P. Wit

promotiecommissie: Prof. dr. P.H. Dejonckere
Prof. dr. J.L.N. Roodenburg
Prof. dr. F. Zwarts

Aan Heit en Mem

Acknowledgments

This research project was supported by the Linguistic Research Foundation, which is funded by the Netherlands Organization for Scientific Research, NWO.

The study is part of the research program of our department: Communication through Hearing and Speech, which is incorporated in the Sensory Systems Group of the Groningen Graduate School for Behavioral and Cognitive Neurosciences (BCN).

Publication of this thesis was financially supported by:
A.C.M. Ooms Allergie B.V.; ASTA Medica B.V.; Astra Pharmaceutica B.V.; Entermed B.V.; Glaxo Wellcome B.V.; GN Danavox Nederland BV; Hoechst Marion Roussel B.V.; Mediprof Holland B.V.; Nobel Biocare Benelux B.V., leverancier van BAHA™ - het Botverankerd Hoortoestel; Oticon Nederland B.V.; Prof. dr. Eelco Huizinga Stichting; Veenhuis Medical Audio B.V.

Printing:
Drukkerij van Denderen B.V. Groningen

ISBN: 90-9009901-8

Contents

1

Introduction

Voice production is a prerequisite to lead a normal live, as it makes social interaction with the environment possible. Generation of vocal sound enables humans to communicate very explicitly. Modification of vocal sound by the articulatory organ results in speech, which carries information from speaker to listener(s). Apart from the content, this information is given additional meaning by varying intensity, pitch and timbre in specific ways, thus creating an emotional context. Spoken language thereby exceeds all other forms of communication.

Vocal sound is based on the generation of pressure differences in the larynx. These pressure differences are the result of vibratory actions of the vocal folds on the expiratory air flow from the lungs. The cyclical opening and closing of the vocal folds is influenced by physiological and aerodynamic factors, such as subglottal pressure, vocal fold tension, and elastic and Bernoulli forces (cf. 1-3). The precise and complex interaction of these variables in phonation is an important topic for study.

Thanks to numerous investigations a fairly good picture now exists of basic aspects of phonation. Information on what might be regarded as an ideal example of the male voice has been implemented in a number of models (2,4,5). The models have a physiological basis; one might expect that all kinds of voices can be described with these models. However, results acquired while utilizing these models indicate an abstraction of the processes that take place during phonation. The shortcomings of the models are diverse. First of all, all models use a limited number of variables, which is inherent in their application to make processes controllable. A correct representation of processes, especially if these are non-linear, such as phonatory processes, is difficult to realize. Another shortcoming deals with the constraints that variables in the models should apply. Information on these constraints is rather sparse. Data from voice research might be used to supply this valuable information. A last shortcoming of models is the difficulty that exists in adequately describing the relation between physiology and perceptive characteristics of speech. A close examination of both physiological and perceptive intra-individual data might improve the development of more accurate models.

The absence of hard data on the variation of voice physiology and voice quality characteristics plays an important role in speech technology. On the one hand, in the field of speech synthesis it is not clear how certain voice types are to be realized, whereas in the field of speech recognition, on the other hand, the variability of voice quality, as well as the incapability to make a distinction between voice and articulation, present obstacles in the transition of speaker dependent equipment to products that can recognize speech of large groups of speakers.

Thus, both with respect to physiological as well as quality features of voices information is lacking: the variation in related characteristics is unknown, which presents a hindrance in several ways. The deficiency of a clear frame of reference of what should be regarded as "normal" or "average" in different populations, creates an obstacle in generalizing results of investigations from one group to another. Classification of groups might be: "extremely well - good - normal - sub-normal - pathological voicing", "male - female - children voices", "good - normal - bad voice quality" or "normal - defective constitution of the vocal apparatus".

A few examples are given to illustrate that a deficiency of a frame of reference with respect to voice quality presents a disadvantageous situation for current research: voice quality is an important factor in sociolinguistic research (cf. 6,7). However, there are insufficient data on voice and speech quality features in large groups to differentiate between normal and abnormal characteristics. It is therefore difficult to relate speech quality directly to social status. Another example: it is difficult to differentiate between normal and pathological voicing when a patients visits a doctor with vague voice complaints. This differentiation is also difficult to make for some physiological voice characteristics (8). A similar problem emerges with voice entrance examinations, performed to decide about the admittance of candidates to specific schools or studies, preparing students for a profession with high vocal demands (speech therapy/logopedics, theater, conservatory, teacher). The lack of data with information on the variation of vocal characteristics might lead to radical decisions, which are not always as solid as they should be, regarding the implication of the decision for the candidate. Besides indirect laryngoscopy performed with mirror examination, additional investigatory methods, such as ascertaining vocal physiology and evaluation of voice quality characteristics should be used to give a more firm judgment on the robustness of the vocal apparatus.

The aim of this study is to assess information on voice quality features and to ascertain the variability of these features in specified groups. Groups were created based on gender and status of vocal training, in order to study the influence of these grouping variables on selected voice quality features.

Gender was chosen as a grouping variable, because previous investigations clearly demonstrated differences in voice quality characteristics between men and women. These differences have implications for the creation of a normative database, concerning its proposed function as a frame of reference. Vocal training was intentionally introduced to give direction to what might be regarded as good vocal characteristics, as compared to characteristics of subjects without vocal training.

Characteristics of the vocal apparatus and voice quality features can be acquired in many ways. Four practicable methods, easily employed in a clinical environment and extensively outlining the vocal apparatus and voice function, are used in this study. Results of these investigations are described in the following chapters.

Chapters 2 and 3 give a close description of the generator of vocal sound, that is, the larynx with vibrating vocal folds. Videolaryngostroboscopy offers images that can be used to concentrate on both aspects of laryngeal appearance, as well as features of glottal functioning. Chapter 2 gives the results of standardized evaluation of laryngeal appearance and glottal functioning. Chapter 3 focuses on a specific feature of glottal functioning, namely glottal closure. Glottal closure has an important influence on the quality of the generated speech signal and it is associated with perceived breathiness. The clinical relevance of evaluating glottal closure is based on the relation with robustness of the larynx, that is, the resistance to vocal complaints during voice demanding tasks.

Vibrational movements of vocal folds result in modulation of air flow and generation of pressure differences. The glottal volume velocity waveform (GVVW) can therefore be regarded as the information carrier of the voice source. Chapter 4 gives the results of an extensive study on voice physiology, employing GVVWs that were acquired with the so-called Rothenberg mask.

Susceptibility to vocal fatigue is the topic of chapter 5. Vocal fatigue is related to specific physiological mechanisms taking place while increasing sound intensity. Groups with differing degrees of susceptibility are compared regarding the underlying physiology of phonation.

Pressure differences created at the glottal level are modified in the vocal tract. These articulatory processes give meaning to the basic voice source signal, and the resultant product is speech. In chapter 6 speech of large groups of subjects is perceptually evaluated with a carefully developed standardized scaling instrument. Results reflect differences in underlying phonatory and articulatory mechanisms.

The pressure differences generated by the voice source can be quantified as sound pressure level (SPL). SPL can be varied along a range from the softest to the loudest possible phonation. Each individual has its own range. The magnitude of the range along a persons also individually differing frequency range is of importance, because it gives information on the possibilities and limitations a voice has during speech production in all of its aspects (from soft speech to shouting). Measured ranges also give information on the quality of the voice source, that is, pathological processes of the vocal apparatus can have a specific manifestation on the produced intensity and frequency ranges. The phonetogram gives a two-dimensional representation of these individual intensity and frequency ranges and is therefore of clinical importance. The evaluation of an individual phonetogram, however, presents a problem. It is difficult to compare two-dimensional data with reference values. Chapter 7 offers a new method to evaluate phonetograms.

In chapter 8 this new method is used along with a more commonly used method, to analyze voice capabilities of large groups of subjects. Normative data will be given and differences between groups are also presented.

In this study inter-individual differences in vocal characteristics are determined between subjects without and subjects with vocal training. Chapter 9 focuses on the effect of vocal training on intra-individual differences in both phonetograms features and phonation times.

Finally, in chapter 10 the results of the investigations presented in the previous chapters are summarized and conclusions are drawn.

References

1. van den Berg Jw, Zantema J, Doornenbal P. On the air resistance and the Bernoulli effect of the human larynx. J. Acoust. Soc. Am. 1957;29:626-631.
2. Titze I, Talkin D. A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. J. Acoust. Soc. Am. 1979; 66: 60-74.
3. Story B, Titze I. Voice simulation with a body-cover model of the vocal folds. J. Acoust. Soc. Am. 1995;97:1249-1260.
4. Ishizaka K, Flanagan J. Synthesis of voiced sounds from a two-mass model of the vocal cords. Bell Sys. Tech. J. 1972;51(6):1233-1268.
5. Cranen B. The acoustic impedance of the glottis. Thesis. K U Nijmegen, 1987.
6. Bezooijen R van. De attributie van persoonlijkheidskenmerken op grond van spraak. In: Creten J, Geerts G, Jaspaerts K, eds. Werk in uitvoering. Momentopname van de sociolinguistiek in België en Nederland. Leuven/Amersfoort: Acco, 1986:13-26.
7. Bezooijen R van, Boves L. The effects of low pass filtering and random splicing on the perception of speech. J. of Psycholinguistic Research 1986;15:403-427.
8. Schutte H. The efficiency of voice production. Thesis. Groningen, 1980.

2

Standardized Laryngeal Videostroboscopic Rating: Differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency and age

Arend M. Sulter, Harm K. Schutte[*] and Donald G. Miller[*]

Department of Otorhinolaryngology, University Hospital Groningen and
[*]Voice Research Lab, Department of Medical Physiology, Faculty of Medicine, University of Groningen, Groningen, the Netherlands

Previously published in the *Journal of Voice* 1996,10(2),175-189.

INTRODUCTION

Information on vocal structures and function can be acquired in many ways (1). Among acoustic, aerodynamic, perceptual and physiologic investigatory methods, all focusing on products and derivatives of the voice source, laryngoscopy can be used to examine the bodily source of vocal sound, that is, the larynx with vibrating vocal folds. The visual study of the dynamic aspects of the glottis during voice production can yield information of both clinical and scientific value (2-8).

The normal behavior of vocal folds vibrating during voice production is a matter of continuing investigation (9). Assessing information on normal voice production is essential to determine fundamental principles and mechanisms of voice function and to formulate criteria which can be used to detect abnormal laryngeal appearance and glottal function. Knowledge pertaining to normal laryngeal appearance and glottal function can be based on an large normal database, established after investigating a large group of normal (voice-healthy) subjects. Rating forms have been used to evaluate normal laryngeal anatomy and glottal functioning. Vocal fold closure --that is type, completeness and duration of closure-- showed gender- and age-specific characteristics (10-13), and effects of pitch and intensity on glottal functioning have been determined (14-19).

However, most of the concluding remarks of the studies are based on a limited number of investigated subjects, and in many cases laryngostroboscopy is performed with the less adequate method of flexible laryngoscopy (20,21). Moreover, a number of these studies were not intended to create a normative database and therefore lack the necessary experimental design. These considerations suggest that the observations documented in these studies are in need of further confirmation (10,22-24).

A normative database is helpful to indicate deviant vocal function and to detect potentially pathologic voicing patterns. To clarify what should be regarded as "poor" or "weak" vocal function, laryngeal videostroboscopic data on good voices with a high level of vocal endurance could be used. Data on laryngostroboscopic characteristics of singers reveal differences with respect to untrained subjects, reflecting the acquired or naturally present special laryngeal functioning (25).

In the present study a large group of male and female untrained subjects was investigated laryngostroboscopically with a rigid endoscope in order to generate a normative database on laryngeal behavior during voice production. A large group of amateur singers was also investigated and their laryngeal characteristics were compared with these of the untrained group. The acquired information was used to answer the following research questions:

1. What differences can be found between male and female subjects in specified laryngeal characteristics?
2. Do trained groups differ in laryngeal characteristics compared to untrained groups?

Using information about the influence of selected variables such as intensity, pitch and age on laryngeal characteristics the next questions were answered:

3. What is the influence of intensity on laryngeal characteristics?
4. What is the influence of pitch?
5. Is there a relation between age of the subject and changes in laryngeal characteristics?

METHODS

Subjects

A total of 224 Dutch untrained and trained subjects of both genders, categorized accordingly into 4 groups, were investigated. The untrained subjects were recruited from groups of students and volunteers without vocal complaints or history of vocal pathology. The group consisted of 92 females and 47 males. The mean age for the female subjects in this subgroup was 20.3 years, ranging from 17 to 44 (median 19 years; standard deviation [SD] 7.37), while the mean age for the male subjects was 25.0 years, ranging from 17 to 35 (median 25 years; SD 4.68 years). Eighteen of the female, and 16 of the male subjects were smokers.

42 female and 43 male amateur singers with a minimum of two years of vocal training served as another group. The vocal training could either consist of singing in a choir that organized rehearsals with a minimum frequency of once a week, or receiving individual singing lessons with a similar minimum frequency. All choirs had a professional conductor and used auditions to admit new members. Although a minimum of 2 years of organized singing was used as a selection criterion to be included in the trained group, about 60% of the trained subjects had a considerably longer history of singing in a choir (> 5 years). The mean age of the female trained group was 35.1 years, ranging from 18 to 59 (median 34 years; SD 11.86 years), and the mean age of the male subjects was 47.5 years, ranging from 21 to 75 (median 49 years; SD 18.52 years). Five of the female, and 11 of the male trained subjects were smokers. Because all participants in this study volunteered, we refrained from matching according to age.

Instrumentation

Laryngeal examinations were performed with a Wolf 90° rigid endoscope (Model 4450.57). A Brüel & Kjær 4914 Rhino-Larynx Stroboscope was used for stroboscopic investigation. The endoscope was connected to a Panasonic CCD camera (Model WV-CD 110E). Images were recorded on a Sony Betamax videorecorder SL-C9 ES PAL. All laryngeal videostroboscopic examinations were performed by one of the authors (HKS), a phoniatrician with extensive experience.

Procedure

Subjects were seated in a chair during the examination. Prior to the actual introduction of the endoscope each subject received information about the procedure, making him or her aware of the harmless character of the investigation. To determine the person's control of the voice, the subject was asked to perform a few preliminary tasks. This step increases the chance of a successful completion of the tasks during actual videostroboscopic examination. Topical anaesthesia (Xylocaine®) was administered to all subjects to expedite the examination[1]. The subject was asked to hold a contact microphone against the skin in the neck region, providing an input source for setting the flash rate of the stroboscope. Next, the investigator took the protruded tongue and held it during the examination slightly out of the mouth with a gauze. The endoscope was introduced with a 90° rotation, in order to keep the lens clean, over the midline of the tongue body. The endoscope was then rotated back, once the lens was in the oropharyngeal space, behind the tongue. During this procedure the touching of pharyngeal structures is carefully avoided. The video recording was started with an overview of the hypopharynx and larynx during relaxed breathing of the subject. Before the phonation tasks were begun the investigator focused on the vocal folds. During the tasks, the first part of each task was recorded with continuous light and then consecutively with stroboscopic light in slow motion mode with a predetermined frequency difference (26). After the procedure the recorded images were shown to the subject and information was given on anatomy and vocal function, which helped to motivate the cooperation.

	Intensity		
Pitch	Soft mean (SD)	Normal mean (SD)	Loud mean (SD)
Male			
Low	124.3 (17.31)	120.2 (17.40)	126.6 (19.21)
Normal	172.8 (23.43)	171.6 (26.73)	172.3 (26.24)
High	249.6 (44.63)	258.7 (56.29)	255.8 (62.26)
Female			
Low	187.8 (34.46)	203.6 (26.11)	196.0 (27.13)
Normal	259.1 (36.78)	261.1 (36.45)	262.8 (31.83)

Table 1. Mean frequencies and standard deviations (SD) in Hz of pitches produced during phonatory tasks in male and female subjects.

Phonatory tasks

The tasks consisted of the production of an /i/-like vowel sound on three intensities (comfortable, soft, loud) with three different pitches (comfortable, low, high). The intensities and pitches were chosen by the subject with the investigator's approval. Allowing the subject to chose the pitch and intensity level presumably resulted in a natural comfortable voice production. Information on the pitches produced is given in Table 1. Absolute values for sound pressure levels were not obtained

[1]A study by Peppard et al. (23) showed no influence of topical anaesthesia on vocal function.

due to the automatic gain control. Subjects were encouraged to produce an /i/-like vowel sound, to optimize the view of the larynx by obtaining a maximal anterior position of the cranial part of the epiglottis. Starting at a comfortable intensity and pitch, hereafter referred to as "normal", each subject was asked to produce phonations with relatively soft, followed by relatively loud intensity, repeating this procedure with relatively low and high pitch. Care was taken to avoid transition from chest to falsetto register; however, in a number of female subjects phonation in falsetto voice could not be avoided. Each successful registration of a combination of specific pitch and intensity level resulted in a token.

Rating form

A new form was created for the rating experiment, using elements from forms published previously (16,27,28). The form was designed with a normal larynx in mind; therefore only scales were incorporated denoting variation of laryngeal features within a normal population.

The form contained two parts (see appendix). The first part consisted of scales relating to overall laryngeal anatomy (larynx/pharynx ratio; epiglottal shape; asymmetry in the arytenoid region) and tendency of supraglottal anatomical structures to show compensatory movements during the variation of intensity. These scales had to be rated during a run-through of the registered videomaterial of the subject under investigation. The second part consisted of scales relating to a visual impression of the vocal folds (thickness; width; length; elasticity) and glottal functioning (amplitudes; phase differences; vocal fold closure). This part had to be rated separately for each token within a subject. The appendix shows the rating scales and accompanying instructions.

Rating experiment

The rating experiment took place in a period extending over three months. Each session lasted no longer than two and one half hours to ensure optimal concentration. Three judges, familiar with laryngostroboscopic video registrations because of their almost daily use of videostroboscopy in a clinical and research situation for at least more than three years, observed the acquired material and systematically scored their impressions on the form. In the first session the use of the rating scales were practiced and sufficient agreement was attained after discussion. Sufficient agreement in this case means that at the end of the practice session the three judges used the scale ends in the same way, having a clear image of what the scale represented, and that the score did not substantially differ with more than two points. At the end of the rating experiment the images of 14 subjects with 36 tokens were rated again to provide re-test data, and intra-judge reliability was calculated.

The judges were seated at a distance of 1 m from a television screen (Sony KV M14D) with a diameter of 34 cm. All tokens of one subject were first shown with sound for rating the first part of the form. Then each token was played in slow-motion at 1/10 speed without sound to score the second part of the form[2]. Each judge was given enough time to complete the second part of the form, which sometimes required a rerun of the specific token. No discussions of the test material were allowed during the rating procedure.

The quality of the images was checked during the actual laryngostroboscopic recording. Apart from the investigator using the endoscope, a second one checked whether all vocal structures were properly illuminated and visible. If these conditions were not met, the specific task was performed again. Because of thyroid-cricoid approximation and shift of the petiolus, in some (<5%) low pitched phonations the anterior commissure was not visible in male subjects.

All video registrations had been collected on three betamax tapes. At the beginning of each session one of the three tapes was randomly chosen, starting the tape at the point where it had last been used.

Intensity and pitch level of each token were determined and the fundamental frequency in hertz was determined by a sung imitation of the same pitch, which can be done with high accuracy, and measuring, by means of an electroglottograph with neck electrodes, the frequency of the sung tone.

[2]Södersten could not establish significant differences in rating with or without sound (16).

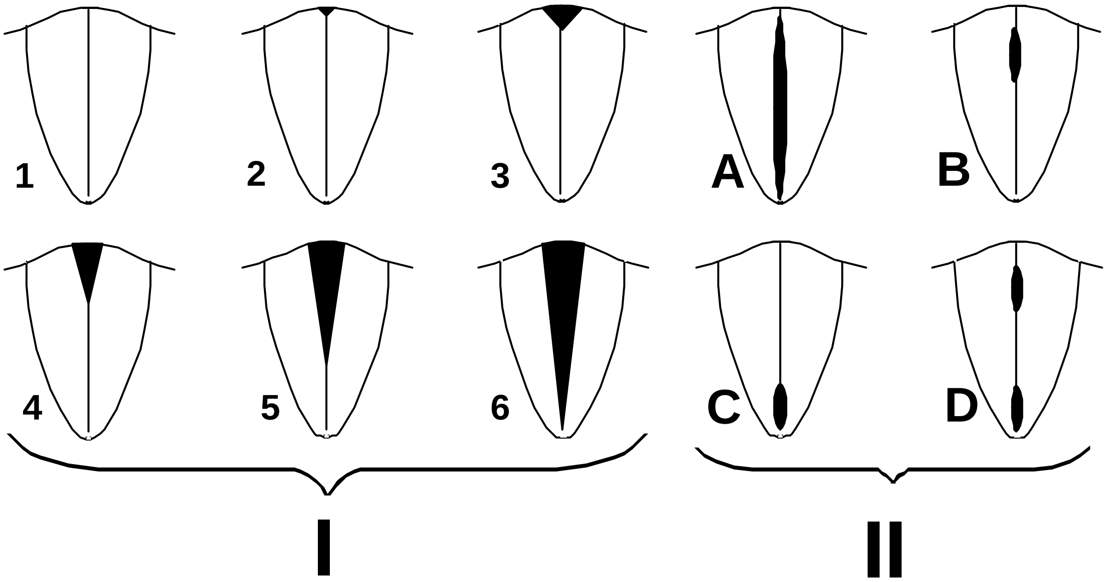


Figure 1. Types of glottal closure (after Södersten et al. Glottal closure and perceived breathiness during phonation in normally speaking subjects. J Speech Hear Res 1990;33:601-11). Category I depicts glottal chinks with increasing open aspect: 1 = complete closure, 2 = incomplete closure in the cartilaginous part, 3 = triangular incomplete closure anterior to the vocal processes, 4 = triangular incomplete closure of the posterior thirds of the folds, 5 = incomplete closure of the posterior two thirds of the folds, 6 = incomplete closure all along the folds. Category II depicts anterior and complex membranous glottal chinks. A = spindle-shaped incomplete closure, closure at the vocal processes, B = spindle-shaped incomplete closure at the posterior thirds of the folds, closure at the vocal processes, C = spindle-shaped incomplete closure at the anterior third of the folds, closure at the vocal processes, D = spindle-shaped incomplete closure at the posterior and the anterior thirds of the folds, closure at the vocal processes and at the middle of the membranous portion.

Data management

Resulting scores were put into a spreadsheet file (Quattro, Borland International Inc.). Ratings from quantitative scales were transformed into an average score. The scale "closure type" (after Södersten, see Figure 1) needed a different treatment in a number of cases. When, instead of supplying three numbers (indicating a closure type 1 to 6), letters (indicating an anterior or complex closure type) were given by the three judges, the modus was taken as the resulting value. In a very limited number of tokens (< 1%) where three different letters or one number were scored, the score of the first author (AMS) was used for further analysis.

The processed spreadsheet was imported into a statistical software package (SPSS, SPSS Inc.) to calculate reliabilities and to generate descriptive and inferential statistics.

Statistical analysis

The original scores of the judges and the re-test scores were used to calculate inter- and intra-judge reliability levels, respectively. The averaged scores on scales were used for descriptive statistics. To determine the effect of factors such as gender, vocal training and intensity level on scales, analysis of variance (ANOVA) (with covariants fundamental frequency and age) and χ^2-tests were performed. The covariant "age" was introduced, since studies had demonstrated the effect of age upon laryngeal appearance and glottal function (10,11,13,29). If a significant interaction was present between factors, separate ANOVA's were performed on each factor level. Because of the many tests performed, a conservative probability level α=0.005 was used with respect to the Bonferroni inequality.

RESULTS

The laryngovideostroboscopic experiment

Token	No.	%	Cum. %
1	214	21.2	21.2
2	203	20.1	41.2
3	168	16.6	57.9
4	123	12.2	70.0
5	105	10.4	80.4
6	82	8.1	88.5
7	58	5.7	94.3
8	36	3.6	97.8
9	22	2.2	100.0
Total	1,011	100.0	100.0

Table 2. Number of tokens per subject

Videostroboscopic examinations resulted in the registration of 1011 tokens that could be used for further analysis. In two subjects no registration could be made due to uncontrollable reflexive pharyngeal movements. From Table 2 it is apparent that only a small minority of the subjects was able to accomplish the whole set of tasks. Moreover, a small number of tokens could not be used because one of the judges did not give a score for a particular scale.

Reliability

Inter-judge and intra-judge reliabilities were determined. Inter-judge reliabilities for scales of both the first and second part of the form were determined using α levels. Cronbach's α ranges from 0, indicating no agreement between judges, to 1, indicating complete agreement between judges. Table 3 summarizes the inter-judge reliability analyses. If a reliability level of 0.6 is taken as a sufficiently high level of agreement among judges (30), only a few scales show poor agreement: larynx/pharynx ratio with α=0.43, thickness of vocal folds with α=0.45, and elasticity of vocal folds with $\alpha=0.28$. Two of the three rated phase differences, vertical and horizontal, show values just below $\alpha=0.6$. The low levels of these last two scales can be attributed to the binomial character of the scales. Especially percentage of closure, type of closure, and location of chink were rated with high agreement. Other studies confirm this observation (9,22,25,31). Since all but one scale showed significance at a level $p<0.0001$, the results of the ratings were used for further analysis; however, caution is advised in drawing conclusions from the first three scales mentioned, which show α levels below 0.45.

	Item	Range	m.i.c.	α
1	Laryngeal Appearance			
	larynx/pharynx ratio	1 - 3	0.20	0.43
	epiglottal shape	1 - 5	0.46	0.68*
	asymmetry arytenoid region	1 - 4	0.51	0.76
2	Compensatory Adjustments	1 - 4	0.48	0.74
3	Vocal Fold Appearance			
	thickness	1 - 5	0.21	0.45
	width	1 - 5	0.43	0.69
	length	1 - 5	0.52	0.77
	elasticity	1 - 5	0.11	0.28
4	Amplitudes	1 - 4	0.50	0.75
5	Vocal Fold Closure			
	duration	1 - 4	0.56	0.79
	percentage	0 - 100	0.82	0.93
	type	1 - 6	0.83	0.94
		A - D		
6	Phase Differences			
	vertical	0 - 1	0.27	0.52
	horizontal	0 - 1	0.32	0.58
	lateral	0 - 1	0.50	0.75
7	Location chink	0 - 2	0.73	0.89

Table 3. Categories of scales and inter-judge reliability levels. The range of the interval scale, the mean inter-judge correlation (m.i.c.) and Cronbach's " are given in columns. All reliabilities are significant at a level $p<0.0001$, except * which has a level $p<0.05$.

Correlation coefficients were calculated to offer intra-judge reliability levels. Table 4 presents the calculated correlation coefficients together with probability levels. Probability levels for calculated correlation coefficients vary between $p<0.1$ and $p<0.001$. Generally, the consistency in rating is at a high level in each of the judges. However, certain scales, such as thickness of vocal folds, compensatory adjustments, and larynx/pharynx ratio, are problematical to judge, because of the lack of a direct reference for measure. On the other hand, aspects of laryngeal appearance, such as epiglottal shape and asymmetry of the arytenoid region, as well as aspects of vocal fold appearance, such as length and elasticity, are rated highly consistently. Especially amplitudes of vocal fold excursion and the scales representing vocal fold closure and phase differences are rated highly consistently with a probability level $p<0.001$.

	Item	c.c.	p-level
1	Laryngeal Appearance		
	Larynx/pharynx ratio	0.51-0.70	p<0.1
	Epiglottal shape	0.74-0.80	p<0.005
	Asymmetry arytenoid region	0.56-0.66	p<0.05
2	Compensatory Adjustments	0.49-0.64	p<0.1
3	Vocal Fold Appearance		
	Thickness	0.28-0.44	p<0.1
	Width	0.45-0.53	p<0.01
	Length	0.68-0.79	p<0.001
	Elasticity	0.68-0.87	p<0.001
4	Amplitudes	0.68-0.77	p<0.001
5	Vocal Fold Closure		
	Duration	0.65-0.82	p<0.001
	Percentage	0.63-0.98	p<0.001
	Type	0.62-0.71	p<0.001
6	Phase Differences		
	Vertical	0.69-0.71	p<0.001
	Horizontal	0.56-1.00	p<0.001
	Lateral	0.37-0.43	p<0.05
7	Location chink	0.76-0.80	p<0.001

Table 4. Categories of scales and intrajudge correlation coefficients (c.c.). The extremes in correlation coefficients are given together with the probability level (p-level) of the weakest correlation. Calculations are based on the two ratings (test - retest) each judge gave on 14 tokens representing scales 1 and 2, and 36 tokens representing scales 3 to 7.

Laryngeal appearance

Table 5 gives the distribution of scores on the scales representing the laryngeal appearance. The female subjects have a smaller larynx/pharynx ratio, a finding confirmed by ANOVA (see Table 6). There are no apparent differences in epiglottal size; however, it is remarkable that only male subjects had positive scores on omega and deviant shaped epiglottises. This finding is highly significant (χ^2-test, p = 0.00003). Asymmetry in the arytenoid region during phonation is a common observation (23,32). In this study almost half of the subjects were rated having a certain degree of asymmetry (see Table 5); however, neither effects of gender and vocal training, nor an influence of age upon the degree of asymmetry could be established (see Table 6).

score	1		2		3		4		5	
scale	♂	♀	♂	♀	♂	♀	♂	♀	♂	♀
Larynx/ pharynx ratio	3 (3.6)	0 (0.0)	73 (8-6.9)	109 (8-6.5)	8 (9.5)	17 (1-3.5)				
Epiglottal shape	5 (6.0)	5 (4.0)	55 (6-5.5)	92 (7-3.0)	13 (1-5.5)	29 (2-3.0)	3 (3.6)	0 (0.0)	8 (9.5)	0 (0.0)
Asymme-try arytenoid region	0 (0.0)	1 (0.8)	12 (14.3)	8 (6.3)	30 (35.7)	48 (38.1)	42 (5-0.0)	69 (54.8)		

Table 5. Distribution of scores for the rated scales of laryngeal appearance. Gender is separately given in columns. The absolute frequency of score and between brackets the relative frequency is given. Larynx/pharynx ratio: 1=large, 2=normal, 3=small; epiglottal shape: 1=large, 2=normal, 3=small, 4=omega, 5=deviant; asymmetry arytenoid region: 1=severe asymmetry, 2=asymmetry, 3=slight asymmetry, 4=no asymmetry.

	Gender		Vocal training		Gender x Training		Age		
	F	p	F	p	F	p	F	p	rrc
Larynx/ pharynx ratio	8.54	0.004 *	0.01	0.919	1.19	0.276	0.19	0.667	-0.001
Epiglottal shape	4.54	0.034	0.04	0.847	0.01	0.946	1.21	0.272	0.004
Arytenoid asymmetry	1.46	0.229	1.00	0.320	2.18	0.142	0.94	0.332	0.003

Table 6. Analysis of covariance summary table with effects of factors and covariant on scales, representing aspects of laryngeal appearance. Note. df=1, 208 for gender, vocal training, and gender x training. rrc = raw regression coefficient. *$p < 0.005$

	Score							
	1		2		3		4	
Scale	♂	♀	♂	♀	♂	♀	♂	♀
Compensatory adjustments	1 (1.2)	0 (0.0)	6 (7.1)	1 (0.8)	14 (16.7)	12 (9.5)	63 (75.0)	113 (89.7)

Table 7. Distribution of scores for compensatory adjustments. Absolute frequency and, between brackets, relative frequency are given. 1=clearly visible, 2=visible, 3=almost absent, 4=not visible.

Compensatory adjustments

Many subjects show movements of supralaryngeal structures while changing intensity and frequency. The degree of these compensatory adjustments were rated on a 4- point scale (Table 7). Statistical analysis showed a significant relation with the factor gender (see Table 6), male subjects more frequently revealing compensatory adjustments. Though not significant, the covariant age has a low probability level (p = 0.007) with a negative raw regression coefficient, indicating that older persons tend to show more compensatory adjustments.

	Vocal training		Intensity level		Frequency level		Age		
	F	p	F	p	F	p	F	p	rrc
♀	4.62	0.032	1.80	0.167	567.81	<0.001 *	18.78	<0.001 *	-0.760
♂	0.17	0.682	0.17	0.842	491.65	<0.001 *	5.43	0.020	-0.229

Table 8. Analysis of covariance summary table with effects of factors and covariant age on fundamental
frequency, separately presented by gender.
Note. Female subjects (&): df=1, 523 for gender and vocal training. df=2, 522 for intensity level.
Male subjects (%): df=1, 483 for gender and vocal training. df=2, 482 for intensity level.
rrc = raw regression coefficient. *p < 0.005

Fundamental frequency

During the accomplishment of tasks the subjects phonated at freely chosen pitches. Phonations in falsetto register were avoided, whenever possible. Table 1 gives the averaged fundamental frequencies and standard deviations for both female and male subjects. To analyze the influence of the factors gender, vocal training, age, intensity level and frequency level on fundamental frequency, a four-way ANOVA was performed with age as covariant. A significant influence of the factor gender on fundamental frequency was established (F(1,1008) 1280.36, p<0.001). Because a significant interaction between gender and frequency level had been found (F(2,1007) 24.88, p < 0.001), separate ANOVA's were performed hereafter for both male and female subjects to

	Gender		Vocal training		Intensity level		Interaction
	F	p	F	p	F	p	
Appearance thickness	13.66	<0.001 *	0.08	0.781	0.74	0.479	
width	47.16	<0.001 *	0.44	0.509	0.47	0.628	
length	238.06	<0.001 *	2.42	0.120	3.14	0.044	
elasticity	53.45	<0.001 *	2.20	0.139	4.50	0.011	
Amplitudes	40.50	<0.001 *	0.55	0.461	143.81	<0.001 *	
Closure duration	38.04	<0.001 *	0.94	0.334	181.87	<0.001 *	(a)
♀			2.57	0.110	83.52	<0.001 *	A1
♂			0.22	0.639	95.62	<0.001 *	A2
percentage	18.32	<0.001 *	0.23	0.628	172.60	<0.001 *	(b)
♀			0.33	0.564	116.46	<0.001 *	B1
♂			0.10	0.751	63.28	<0.001 *	B2
type	47.09	<0.001 *	0.11	0.742	116.14	<0.001 *	(b)
♀			0.73	0.393	83.40	<0.001 *	B3

Table 9. Analysis of covariance summary table with effects of factors on scales, representing aspects of vocal fold appearance and function.

Note. df=1, 1009 for gender, vocal training, and gender x training. df=2, 1008 for intensity level, gender x intensity level, training x intensity level, and gender x training x intensity level.
(a) interaction between gender and training (p = 0.004);
(b) interaction between gender and intensity level (p < 0.001)
A1 df=1, 524 for vocal training. df=2, 523 for intensity level, and training x intensity.
A2 df=1, 483 for vocal training. df=2, 482 for intensity level, and training x intensity.
B1 df=1, 523 for vocal training. df=2, 522 for intensity level, and training x intensity.
B2 df=1, 483 for vocal training. df=2, 482 for intensity level, and training x intensity.
B3 df=1, 472 for vocal training. df=2, 471 for intensity level, and training x intensity.
B4 df=1, 387 for vocal training. df=2, 386 for intensity level, and training x intensity
*p < 0.005

determine the effects of the remaining factors vocal training, intensity level and frequency level. Table 8 summarizes the results of these analyses. A significant

	Fundamental Frequency			Age		
Vocal Fold	F	p	rrc	F	p	rrc
Appearance thickness	73.23	<0.001*	-0.002	5.86	0.016	-0.003
width	28.32	<0.001*	-0.001	13.47	<0.001*	-0.005
length	203.54	<0.001*	0.003	0.38	0.538	0.001
elasticity	23.57	<0.001*	-0.001	8.56	0.004*	0.003
Amplitudes	116.28	<0.001*	0.002	23.60	<0.001*	0.005
Closure duration	70.28	<0.001*	0.002	0.33	0.566	0.001
♀	70.80	<0.001*	0.002	9.00	0.003*	0.006
♂	37.91	<0.001*	0.003	6.88	0.009	-0.004
percentage	78.35	<0.001*	-0.052	23.11	<0.001*	0.175
♀	15.67	<0.001*	-0.037	3.09	0.080	0.120
♂	5.48	0.020	-0.026	15.87	<0.001*	0.165
type	58.21	<0.001*	0.003	49.01	<0.001*	-0.018
♀	11.97	0.001*	0.002	13.40	<0.001*	-0.017
♂	0.03	0.855	-0.000	24.67	<0.001*	-0.012

Table 10. Analysis of covariance summary table for covariants. Scales with interaction between factors are separately treated by gender. See table 9 for specific degrees of freedom. rrc = raw regression coefficient. $^*p < 0.005$

effect of the factor frequency level was found. Post hoc LSD tests with a significance level p=0.005 showed significant differences among all of the three frequency levels in both male and female subjects. Age had an influence on pitch in both male and female subjects, but this was only significant in females. The raw regression coefficient had a negative value in both cases, implying a decrease in pitch with increasing age.

Vocal fold appearance

Three-way ANOVA's with covariants fundamental frequency and age were used to determine the effects of the factors gender, vocal training and intensity level on the scales thickness, width, length and elasticity. Table 9 and 10 summarize the results. Gender had a significant effect on all four scales, with female subjects having thinner, wider, shorter and slacker vocal folds (see Table 9). No other statistically significant effects could be established, but a positive effect of the intensity level on length and elasticity was found (p=0.044 and p=0.011, respectively). Generally, with increasing intensity length decreases and elasticity increases. Fundamental frequency had a significant influence on all four scales (see Table 10). With increasing frequency vocal folds are rated thinner, narrower, longer and more tense. Age also had a significant effect on the scales width and elasticity. With increasing age vocal folds are rated narrower and slacker. Thickness was not significantly affected by age; however, a low probability was found (p = 0.016) with a negative raw regression coefficient, implying decreasing thickness with increasing age.

Amplitudes

Quantitative judgments about the excursions of vibrating vocal folds were given. With ANOVA significant effects of the factors gender and intensity level were established. Female subjects were rated as having larger amplitudes than males. Amplitudes are rated significantly larger with increasing intensity (post hoc LSD, p=0.005) (see Table 9). Fundamental frequency and age have a significant influence on amplitude. An increase of both these covariants results in smaller amplitudes (see Table 10).

Vocal fold closure

The scale "closure type" consists of several glottal configurations to be found in the most closed phase of the glottal cycle (see Figure 1). These configurations can be divided into two categories: the numbers 1 to 6 represent a glottal configuration with a gradually increasing closed aspect and are hereafter referred to as category I, whereas the configurations depicted with letters A to D present deviant closure types with glottal gaps present in the membranous or anterior part of the glottis. This latter category hereafter will be referred to as category II. Because of the different character of the two categories separate statistical analysis was performed for each. Three-way ANOVA presented significant interactions in the three scales representing aspects of vocal fold closure. For the scale duration, an interaction was found between gender and training, and for the scales percentage and type, there was an interaction between gender and intensity level. Therefore two-way

ANOVA's were also performed separately for male and female subjects. Table 9 summarizes the results of the performed ANOVA's with covariants. The factor gender showed a significant effect on the scales duration, percentage and closure type of category I, with male subjects having a shorter closure, a higher percentage of closure and an even more closed type. In both male and female subjects significant effects were found for the factor intensity level regarding the scales duration, percentage and type of closure (category I). Post hoc LSD tests indicated significant differences among intensity levels with increasing closure for louder intensities. The covariant fundamental frequency had a significant effect on duration, percentage of closure and closure type (category I). With increasing frequency, closure is briefer, and in the female subjects a present glottal gap becomes larger. Depending on gender, age had also a significant effect on the scales representing vocal fold closure. In female subjects the duration of closure becomes briefer with increasing age, while the opposite appears to happen in male subjects; however, this last effect is not significant (p=0.009). With increasing age all subjects showed a more closed glottal configuration, and in male subjects the percentage of closure increased.

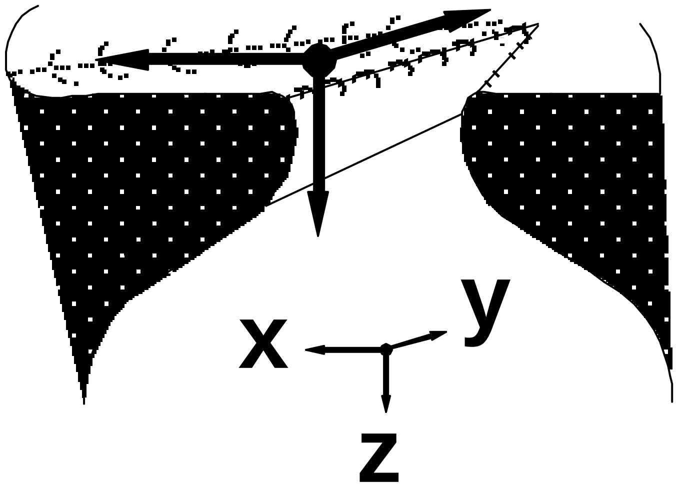


Figure 2. Coronal cross-section of vocal folds with axes delineating directions of possible phase differences: x-axis, lateral phase difference; y-axis, horizontal phase difference; and z-axis, vertical phase difference.

With χ^2 tests differences in closure category (I or II) were analyzed (see Figure 1). Male subjects significantly more often presented closure types from category II (χ^2=22.12, p<0.00001). Gender differences were revealed most significantly by the frequency of type C closure (anterior gap) in male subjects (χ^2=28.10, p<0.00001). Trained subjects had no significant differences in closure type.

Phase differences

During vocal fold vibration, differences between parts of the vocal fold in the cycle of opening and closing can exist. If there is a difference in the cycle between caudal and cranial parts of the vocal folds, this phenomenon is called a vertical phase difference, and in the case of anterior-posterior difference it is called a horizontal phase difference. A difference in phase between vocal folds is called a lateral phase difference. Figure 2 presents these phenomena schematically.

Male subjects were rated more often with a vertical phase difference (χ^2=17.58, p=0.00003), whereas a horizontal phase difference was found more frequently in female subjects (χ^2=33.70, p<0.00001). Generally, a lateral phase difference was observed as frequently in male as in female subjects. Subjects with vocal training showed no differences with respect to a vertical and a horizontal phase difference; however, a lateral phase difference was observed more frequently in this group (χ^2=26.33, p<0.00001).

Chink location

An incomplete glottal closure during the most closed part of the glottal cycle is a common finding (9). The location of this glottal chink was designated, according to appearance, as membranous, cartilaginous, or not visible. In the last case the vocal folds are assumed to close completely.

Membranous and cartilaginous glottal chinks were observed more frequently in female subjects, while male subjects more often received the rating "not visible" (χ^2=87.5, p<0.00001). Trained subjects also received the rating "not visible", and this difference with untrained subjects is significant (χ^2=12.8, p=0.00167).

DISCUSSION

Standardized rating of aspects of laryngeal function has already revealed many characteristics of the voice source, as well as effects of varying sound intensity and pitch. Because no quantitative measurements are made, the validity of conclusions from these studies depends strongly on agreement among the judges, as well as on adequate definitions of aspects to be rated. Reliabilities of rated scales in this study generally were high, with values well above Cronbach's α=0.65. The exceptions to these are the scales thickness and elasticity of vocal folds, and the larynx/pharynx ratio. Thus conclusions regarding these three aspects are to be treated with caution. Reliabilities of scales representing vocal fold closure were close to figures presented in previously published articles (11,15,22,23,25,31,33,34). The various aspects are treated separately below.

Fundamental frequency

The fundamental frequency had a significant influence on almost all scales representing vocal fold appearance and glottal functioning. In this study the values of fundamental frequencies, averaged for each condition, fell within the 20 to 50% and the 10 to 50% absolute frequency range of the male and female groups, respectively (35). This part of the frequency range corresponds to the continuum of fundamental frequencies used in Dutch speech (36) and therefore offers a good representation of the voice source normally used in speech. In both genders the lowest fundamental frequencies in this study are a few semitones above the reported lowest frequency in the Tielen study. The fact of sustaining of a tone, as well as the protruded tongue position might explain the higher fundamental frequency of subjects, when asked to phonate at a comfortable pitch (see also Södersten) (37). Using flexible fiberscopic examination, Pemberton et al. (33) found pitches (normal frequency level) of 218 and 128 Hz for female and male subjects, respectively. These values are lower than the pitches in our study.

Laryngeal appearance and compensatory adjustments

Each rating of a subject's laryngostroboscopic images started with judging laryngeal appearance and compensatory adjustments. Although only a few differences between groups were found in these scales, the diversity in laryngeal appearance is apparent in the summarized ratings (see Tables 5 and 7), specifically with respect to asymmetry in the arytenoid region and epiglottal shape. The larynx/pharynx ratio relates to the space the larynx occupies within the hypopharynx. A large part of the hypopharynx consists of the sinus piriformis cavities, which might play an acoustic role in the singing voice (38). During the act of singing the vocal tract undergoes spatial changes (39) and the volume of the sinus piriformis cavities may be affected in many ways (32). However, in executing the phonatory tasks of this study no differences could be established between trained and untrained subjects for the larynx/pharynx ratio.

Asymmetry in the arytenoid region is in most cases based on the many positions the cuneiform cartilages can assume in the larynx within the aryepiglottic folds and atop the arytenoids. In the laryngostroboscopic view the cuneiform cartilage is the most obvious structure in the posterior laryngeal region (32).

The presence of deviant and omega shaped epiglottises exclusively in male subjects was an unexpected finding, and no relevant literature on this topic is known to the authors.

Vocal fold appearance

Studies dealing with aspects of vocal fold appearance have revealed differences between gender, effects of varying intensity and fundamental frequency, and the influence of age.

In his laminagraphic study Hollien (40) measured vocal fold thickness in male and female subjects and determined the influence of fundamental frequency. Our results are in agreement with his measurements, which revealed that males have thicker vocal folds than females, and that increasing frequency is related to decreasing thickness of vocal folds. This last observation can be explained by the volumetric principle, according to which the volume of vocal fold tissue is constant. A model proposed by Titze et al. (41) includes a minor effect of decreasing thickness on increasing frequency. Honjo (13) investigated the effect of age on vocal fold appearance and described two possible major changes with age: women more often show edema, while vocal fold atrophy can be found more frequently in men. In this study no significant change in thickness was established for age.

In this study the female subjects had a higher averaged rating than males for the width of the vocal folds. This might be due to the fact that in absence of absolute scale, the appearance of width is related to the appearance of length, the shorter vocal folds of women suggesting greater width. The absence of absolute scale might also explain the low level of reliability of this item. Width also correlated negatively with fundamental frequency and age. In their model Titze et al. (41) predicted a decrease in fundamental frequency with an increase in width, which is in accordance with the results of the present study. The decreasing width with age might be related to the increase in size of vestibular folds, reported by Ferrerri (42). The more medially prominent vestibular folds obscure the vision at the lateral parts of the vocal folds, resulting in a smaller rating for width. However, in their study Gracco et al. (29) could not confirm the observation of Ferrerri.

The length of the vocal folds has been a topic of many investigations, and after Farnsworth (43) other investigators also established an increasing length with increasing frequency, both in vivo (44) and, with models, in vitro (41,45). The linearity of this relation, however, is still debated (46). The observation in this study, that males have longer vocal folds than females is supported by anatomical studies (47-49).

The visual impression of the elasticity of vocal folds is generally assumed to be based on the presence of mucosal waves on the surface of the vocal folds. Conditions for appearance of mucosal waves include interaction between body and cover of the vocal folds (27,50,51), stress in various tissue layers (19,41), and the degree of hydration of the vocal fold (29,41). The low reliability level for the scale elasticity showed that the judges differed in rating this scale. To improve the reliability, a more adequate description of elasticity should be elaborated.

In agreement with a study by Bless et al. (28), reporting a greater mucosal wave in females, this study also found higher ratings on elasticity for female subjects compared to male subjects. The decreasing elasticity with increasing fundamental frequency is probably based on stiffening of the cover (19). The apparent increased elasticity observed with aging is difficult to explain, as the senescence of vocal folds is histologically different in men and women. An increase in stiffness of the cover, by deterioration of the mucosal glands (in the ventricular folds) in aged persons, as suggested by Gracco et al. (29), was not confirmed.

Amplitudes

The excursions of the vocal folds during vibration (amplitudes) are determined by subglottal pressure and tension in the vocal folds, as aerodynamic and myoelastic factors, respectively. The increase of amplitudes with increasing sound intensity and decreasing fundamental frequency agrees with effects established in many other studies (2,18,22,52-54). The larger amplitudes in female subjects in this study were established after standardizing with the covariant fundamental frequency. Without this standardization no difference was found. In contrast to the increase in amplitudes with age described by Biever et al. (10), in this study a decrease with age was established. This might result from a change in connective tissue of the body of the vocal folds, producing increased tension in the vocal fold, thereby reducing the magnitude of the excursions.

Vocal fold closure

Adductory forces, comprising myoelastic-aerodynamic factors (55), determine events of the glottal cycle. Vocal fold functioning, especially the glottal configuration during the most closed phase, has been investigated in many studies and gives information on underlying vocal mechanisms.

In a group of 20 adults Bless et al. (28) reported a longer closed phase in male subjects; however, the opposite was found in this study after standardizing for the covariant fundamental frequency. Without the standardization no difference between genders could be established. The observed increase in duration with increasing sound intensity and decreasing frequency was in agreement with results from available studies (18,43,52). In addition, with increasing age in women a decrease in duration of closure was established. Less resistance to subglottal pressure because of atrophy of the thyroarytenoid muscle has been offered as an explanation for this finding (42). Visual observation of duration of closure, however, remains a rather subjective analytic method, and a more accurate determination can be performed with glottography or measuring glottal air flow.

Recently much information has become available about degree and type of closure. In general, men have a more complete glottal closure than women (13,16). The location of a glottal chink is mostly posterior in women (10,12,16), while men show more midmembranous and anterior chinks (13). With increasing age the location of a chink assumes a more anterior position along the vocal folds (10,11). Linville (11) suggested atrophy of either the thyroarytenoid muscle or connective tissue as the cause of this changing appearance of chinks. An increase in sound intensity is positively correlated with glottal closure (9), whereas in most cases an increase in fundamental frequency is related to a decrease in closure (16,22). Our results are all in agreement with the summarized results from these pertinent articles. Although only significant in male subjects, increasing age was related to an increase in completeness of closure. A possible speculative explanation for this is that atrophy of vocal fold structures might force elderly subjects to close the vocal folds more tightly. Men seem to have less difficulty in closing the posterior part of the glottis. This advantage might have an anatomical base, with males having a smaller angle between vocal folds in the resting position (34), and the thicker and longer vocal folds. Thicker and longer vocal folds in male subjects may also be related to vertical phase differences, which will be discussed in the following section. In males the most frequent place for a chink to appear is the anterior part of the vocal folds.

Phase differences

Phase differences were established for the factors gender and vocal training with χ^2-tests, without standardization for fundamental frequency and age. Horizontal and vertical phase differences were already described by Farnsworth in 1940 (43). In this study female subjects more often showed a horizontal phase difference compared to male subjects. This zipper-like closing of the vocal folds might be related to the triangular glottal configuration in women who have a larger angle between the vocal folds. Closure starts at the anterior part, where a critical aerodynamical velocity is exceeded (56), and by the aerodynamical adductory forces more posterior parts of the vocal folds come closer together and thereby also exceed the critical boundary. A vertical phase difference was observed more frequently in male subjects. Thicker vocal folds are more favorable for the development of a vertical phase difference and there is a correlation between vocal fold thickness and fundamental frequency, which could explain the absence of the observation of vertical phase differences at high frequencies in both genders.

The more frequent observation of a lateral phase difference in trained subjects compared to untrained subjects was a surprising finding. Because of age differences in the investigated groups, a possible influence of age was suspected, as was suggested by Biever et al. (10); however, no significant age differences between groups with and without a lateral phase difference could be established.

Chink location

The difference in chink location between males and females is not surprising, as the better closure in males was already established for the scales closure type and percentage. Without standardization for age, trained subjects more often received the rating "not visible".

CONCLUSIONS

The large number of subjects that were investigated in this study produced a database which can be used as a frame of reference for individual laryngostroboscopic images, as well as to study laryngeal behavior during phonation.

The majority of the proposed scales characterizing laryngeal appearance and glottal functioning showed a high reliability level, indicating a high degree of concordance between judges. Only in those scales lacking a direct reference for measure, such as larynx/pharynx ratio and vocal fold thickness, was a lower level of reliability found. The scale "elasticity of the vocal folds" showed a high intra-judge and a low inter-judge reliability. This means that each judge had a clear representation for himself of what is meant with the description elasticity; however, this definition differed among the judges. The implication is that a clearer description of elasticity, which is related to the clinically important mucosal waves, could improve the usefulness of the scale. The generally high level of agreement among the judges demonstrates the feasibility of using standardized rating scales. This should encourage clinicians to adopt a similar procedure of laryngeal assessment for evaluation of the effect of voice therapy or surgery, or for comparing individual laryngeal features to a large database.

Minor differences were established between untrained and trained groups: trained subjects more frequently had a complete glottal closure and showed, surprisingly, asymmetrical vocal fold excursions (lateral phase difference). These minor differences, however, imply that the vocal apparatus basically does not differ between the two groups, which might be explained by the relatively low level of training in the "trained" group. Differences in vocal capacities, as established between the same untrained and trained groups by phonetography in another study (35), seem therefore to be based on a better control over the voice source in trained subjects.

Large differences in laryngeal appearance and glottal functioning were established between male and female subjects. Because analyses of differences were performed with covariant fundamental frequency, this implies that a separate evaluation of laryngostroboscopic images has to be performed with respect to gender, and that female laryngeal and glottal characteristics are not simply comparable to male characteristics transposed by one octave.

Ageing was reflected in specific changes which are characterized by the scales width, elasticity, amplitudes of excursions, and duration of closure. However, the unbalanced age distribution in the different groups may weaken this conclusion.

Evaluation of laryngostroboscopic images ought to be made with due consideration of frequency and intensity level, because these factors have a large influence on many of the scales used for laryngeal assessment. Laryngostroboscopic investigation at one frequency and intensity level provides an image suitable to detect organic vocal fold pathology; however, for a more complete description of the larynx and of glottal functioning, it is preferable to observe a variety of intensity and frequency conditions.

The results of this study are generally in agreement with results from previously published studies; however, a number of our findings have not been reported before. These include the following: compensatory adjustments, anterior type of incomplete glottal closure, and deviant and omega shaped epiglottises are found more frequently in male subjects. With increasing age a more complete glottal closure was found. Horizontal phase differences were seen more frequently in female subjects, whereas vertical phase differences are more often observed in male subjects. These findings might serve as points for further research.

References

1. Baken RJ. Clinical measurement of speech and voice. Boston: College-Hill Press, 1987.
2. Kitzing P. Stroboscopy : a pertinent laryngological examination. J Otolaryngol 1985;14:151-7.
3. Sataloff RT, Spiegel JR, Caroll LM, et al. Strobolaryngoscopy in professional voice users: Results and clinical value. J Voice 1987;4:359-64.
4. Cantarella G. Value of flexible videolaryngoscopy in the study of laryngeal morphology and functions. J Voice 1988;1:353-8.
5. Faure M-A, Müller A. Stroboscopy. J Voice 1992;6:139-48.
6. Sessions RB, Martin GF, Miller SD, Solomon BI. Videolaryngostroboscopy for evaluation of laryngeal disorders. In: Blitzer A, Brin, MF, Sasaki CT et al., eds. Neurologic disorders of the larynx, New York: Thieme 1992:140-6.
7. Teachey JC, Kahane JC, Beckford NS. Vocal mechanisms in untrained professional singers. J Voice 1991; 5:51-6.

8. Koike Y, Imaizumi S. Objective evaluation of laryngostroboscopic findings. In Fujimura O, ed. Vocal Physiology: Voice Production, Mechanisms and Functions, New York: Raven Press 1988:433-42.
9. Södersten M. Vocal fold closure during phonation. Physiological, perceptual, and acoustic studies. Thesis. Stockholm: Department of Logopedics and Phoniatrics, Huddinge University Hospital 1994.
10. Biever DM, Bless DM. Vibratory characteristics of the vocal folds in young adult and geriatric women. J Voice 1989;3:120-31.
11. Linville SE. Glottal gap configurations in two age groups of women. J Voice 1992;35:1209-15.
12. Linville SE, Skarin BD, Fornatto E. The interrelationship of measures related to vocal function, speech rate, and laryngeal appearance in elderly women. J Speech Hear Res 1989;32:323-30.
13. Honjo I, Isshiki N. Laryngoscopic and voice characteristics of aged persons. Arch Otolaryngol (Chicago) 1980;106:149-50.
14. Timcke R, Von Leden H, Moore GP. Laryngeal vibrations: measurements of the glottic wave: part 1 the normal vibratory cycle. AMA Arch Otolaryngol 1958;68:1- 19.
15. Södersten M, Hertegård S, Hammarberg B. Glottal closure, airflow, and voice quality in middle-aged women as related to changes in loudness. Phoniatric and Logopedic Progress Report 1994;9:3-20.
16. Södersten M, Lindestad P. Glottal closure and perceived breathiness during phonation in normally speaking subjects. J Speech Hear Res 1990; 33:601-11.
17. Södersten M, Lindestad P, Hammarberg B. Vocal fold closure, perceived breathiness, and acoustic characteristics in normal adult speakers. In: Gauffin J, Hammarberg B, eds. Vocal Fold Physiology Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms, San Diego: Singular Publishing Group 1991:217-24.
18. Fex S, Löfqvist AG, Schalén L. Videostroboscopic evaluation of glottal open quotient, related to some acoustic parameters. In: Gauffin J, Hammarberg B, eds. Vocal Fold Physiology Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms, San Diego: Singular Publishing Group 1991:273-8.
19. Titze IR, Jiang JJ, Hsiao T. Measurement of mucosal wave propagation and vertical phase difference in vocal fold vibration. Ann Otol Rhinol Laryngol 1993;102:58-63.
20. Casper JK, Brewer DW, Colton RH. Pitfalls and problems in flexible fiberoptic videolaryngoscopy. J Voice 1988;1:347-52.
21. Yanagisawa E. Fiberscopic and telescopic videolaryngoscopy - a comparative study. In: Baer T, Sasaki C, Harris K, eds. Laryngeal function in phonation and respiration. Boston: College Hill Press, 1987:475-84.
22. Pausewang Gelfer M, Bultemeyer DK. Evaluation of vocal fold vibratory patterns in normal voices. J Voice 1990;4:335-45.
23. Peppard RC, Bless DM. The use of topical anaesthetic in videostroboscopic examination of the larynx. J Voice 1991;5:57-63.
24. Muma JR. The need for replication. J Speech Hear Res 1993;36:927-30.
25. Peppard RC, Bless DM, Milenkovic P. Comparison of young adult singers and nonsingers with vocal nodules. J Voice 1988;2:250-60.
26. Berg J. A Δf-Generator and Movie-Adapter Unit for Laryngo-Stroboscopy. Pract Otorhinolaryngol 1959;21:355-63.
27. Hirano M. Clinical examination of voice. New York: Springer-Verlag 1981.
28. Bless DM, Biever DM, Shaik A. Comparison of vibratory characteristics of young adult males and females. In: Hibi S, Bless D, Hirano M, eds. Proc Int Conf Voice Kurume, Kurume, Japan: Dept of Otolaryngology - Head Neck Surgery, School of Medicine Kurume University 1986:117-8.
29. Gracco C, Kahane J. Age related changes in the vestibular folds of the human larynx: A histomorphometric study. J Voice 1989;3:204-12.
30. Nunnaly J. Psychometric theory. New York: McGraw-Hill Company 1978.

31. Rammage LA, Peppard RC, Bless DM. Aerodynamic, laryngoscopic, and perceptual-acoustic characteristics in dysphonic females with posterior glottal chinks: A retrospective study. J Voice 1992;6:64-78.
32. Painter C. The laryngeal vestibule and voice quality. Arch Otorhinolaryngol (Berlin) 1986;243:329-37.
33. Pemberton C, Russell A, Priestley J, Havas T, Hooper J, Clark P. Characteristics of normal larynges under flexible fiberscopic and stroboscopic examination: an Australian Perspective. J Voice 1993;7:382-9.
34. Södersten M, Hammarberg B. Effects of voice training in normal-speaking women: Videostroboscopic, perceptual and acoustic characteristics. Scand J Log Phon 1993;18:33-42.
35. Sulter AM, Schutte HK, Miller DG. Differences in phonetogram features between male and female subjects with and without vocal training. Manuscript accepted by J Voice.
36. Tielen M. Male and female speech. An experimental study of sex-related voice and pronunciation characteristics. Thesis. Universiteit van Amsterdam 1992:180.
37. Södersten M, Lindestad P. A comparison of vocal fold closure in rigid telescopic and flexible fiberoptic laryngostroboscopy. Acta Otolaryngol (Stockh) 1992; 112:144-50.
38. Sundberg J. The acoustics of the singing voice. Sci Am 1977;236:82-91.
39. Sulter AM, Miller DG, Wolf RF, Schutte HK, Wit HP, Mooyaart EL. On the relation between the dimensions and resonance characteristics of the vocal tract: A study with MRI. Magn Reson Imag 1992;10:365-73.
40. Hollien H. The relationship of vocal fold thickness to absolute fundamental frequency of phonation. Proceedings of the fourth international congress of phonetic sciences Helsinki 1961, Mouton & Co 1962:173-7.
41. Titze IR, Talkin DT. A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. J Acoust Soc Am 1979;66:60- 74.
42. Ferrerri G. Senescence of the larynx. Ital Gen Rev Otorhinolaryngol 1959;1:640-709.
43. Farnsworth D. High speed motion pictures of the human vocal cords. Bell Lab Rec 1940;18:203-8
44. Hollien H, Moore P. Measurements of the vocal folds during changes in pitch. J Speech Hear Res 1960;3:157-65.
45. Titze IR. Physiologic and acoustic differences between male and female voices. J Acoust Soc Am 1989; 85:1699-1707.
46. Nishizawa N, Sawashima M, Yonemoto K. Vocal fold length in vocal pitch change. In: Fujimura O, ed. Vocal physiology: voice production, mechanisms and functions. New York: Raven Press, 1988:75-82.
47. Hirano M. The structure of the vocal folds. In: Stevens K, Hirano M, eds. Vocal Fold Physiology. Tokyo: University of Tokyo, 1983:33-43.
48. Kahane J. A morphological study of the human prepubertal and pubertal larynx. Am J Anatom 1978;151:11-20.
49. Hirano M, Kiyokawa K, Kurita S. Laryngeal muscles and glottic shaping. In: Fujimura O, ed. Vocal Physiology: Voice Production, Mechanisms and Functions, New York: Raven Press 1988:49-66.
50. Hirano M. Morphological structure of the vocal cord as a vibrator and its variations. Folia Phoniat 1974;26:89-94.
51. Hirano M, Kakita Y, Ohmani K, Kurita S. Structure and mechanical properties of the vocal fold. In: Lass N, ed. Speech and Language. Advances in Basic Research and Practice. New York: Academic Press, 1982:271-297.
52. Timcke R, Von Leden H, Moore GP. Laryngeal vibrations: measurements of the glottic wave: part 2 physiologic variations. AMA Arch Otolaryngol 1959;69:438 -44.
53. Hirano M, Tanaka S, Yoshida T. Vibratory behavior of human vocal folds viewed from below. In: Gauffin J, Hammarberg B, eds. Vocal Fold Physiology Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms, San Diego: Singular Publishing Group 1991:1-6.

54. Sonesson B. On the anatomy and vibratory pattern of the human vocal folds. With special reference to a photo-electrical method for studying the vibratory movements. Acta Otolaryngol Suppl. 1960:156.
55. vdBerg Jw. Myoelastic-aerodynamic theory of voice production. J Speech Hear Res 1958;1:227-44.
56. McGowan R. An analogy between the mucosal waves of the vocal folds and wind waves on water. Haskins Laboratories Status Report on Speech Research 1990;101/102:243-9.

3

The effects of frequency and intensity level on glottal closure in normal subjects

Arend M. Sulter and F.W.J. Albers

Department of Otorhinolaryngology, University Hospital Groningen, Groningen, the Netherlands

Accepted for publication by *Clinical Otolaryngology*

INTRODUCTION

Vocal sound is based on the cyclical opening and closing of the vocal folds. In this cycle the vocal folds are closed along a part of their length during a certain period. The degree of glottal closure is associated with voice quality and has a perceptive relationship with breathiness (1). Observations on the degree of glottal closure therefore enable a quantitative judgment of the vocal apparatus.

Quantitative data on glottal closure are needed to establish more adequate and complete information of the voice quality. In this investigation a large group of normal men and women was examined with videolaryngostroboscopy to produce standard values. As sound intensity and, to a lesser degree, frequency have been shown to have an effect on glottal closure, the influence of these variables was also investigated (1). Finally differences between men and women were analyzed.

METHODS

Subjects

To obtain information on glottal closure a group of 92 women and 47 men with neither vocal complaints nor vocal abnormality were examined. The mean age for the women was 20.3 years, ranging from 17 to 44, while the mean age for the men was 25.0 years, ranging from 17 to 35.

Instrumentation

Laryngeal examinations were performed with a Wolf 90° rigid endoscope (Model 4450.57). A Brüel & Kjær 4914 Rhino-Larynx Stroboscope was used for stroboscopic investigation. The endoscope was connected to a Panasonic CCD camera (Model WV-CD 110E). Images were recorded on a Sony Betamax videorecorder SL-C9 ES PAL. All laryngeal videostroboscopic examinations were performed by an experienced phoniatrician.

Procedure

Prior to the actual examination topical anaesthesia (Xylocaine®) was administered to all subjects. The endoscope was introduced into the oropharynx carefully so as not to touch any (oro-)pharyngeal structures to avoid gagging, as well as preventing the lens from smearing. Once focused on the vocal folds the video recording was started and the subject was asked to perform a set of phonatory tasks.

Phonatory tasks

The tasks consisted of the production of an /i/-like vowel sound at three intensities (comfortable, soft, loud) with three different pitches (comfortable, low, high). The intensity and pitch were chosen by the subject with the investigator's approval. Allowing the subject to choose the pitch and intensity level presumably resulted in a naturally comfortable voice production, hereafter referred to as "normal". Subjects were encouraged to produce an /i/-like vowel sound, to optimize the view of the larynx by obtaining a maximal anterior position of the cranial part of the epiglottis. Starting with normal intensity and pitch, each subject was asked to produce sounds with relatively soft, followed by relatively loud intensity, repeating this procedure with relatively low and high pitch. Care was taken to avoid transition from chest to falsetto register; however, in a number of women phonation in a falsetto voice could not be avoided.

Glottal closure rating

Vocal fold closure can be scored as a percentage. Figure 1 gives a schematized larynx with delineated vocal folds and incomplete glottal closure. The percentage of closure in this case is 65%.

Three observers, familiar with laryngostroboscopic video recordings, observed the acquired material and noted a percentage of closure for each recording. Each recording was played in slow-motion at 1/10 speed for observation of the most closed phase of the glottal cycle and to have enough time to complete the scoring.

The intensity and pitch level of each recording were determined and the fundamental frequency in hertz was determined by imitating the pitch produced and reading the output on an electroglottograph frequency counter.

At the end of the rating experiment 36 recordings were rated again to provide re-test data, and an intra-observer reliability was calculated.

The original scores of the judges and the re-test scores were used to calculate inter- and intra-observer reliability levels, respectively. The averaged scores on scales were used for descriptive statistics. To determine the effect of intensity and frequency level, as well as gender, analysis of variance was performed. If a significant interaction was present between factors, separate ANOVA's were performed on each factor level. Probability levels below $p<0.05$ were regarded as significant.

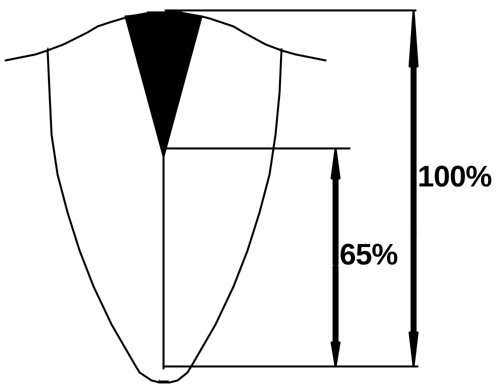


Figure 1. Schematized larynx with delineated vocal folds and incomplete glottal closure. At the top is the arytenoid region and at the bottom the anterior commissure. The black triangle represents the unclosed part of the glottis. Complete glottal closure corresponds to a percentage of 100%. The closure in this figure is 65%.

RESULTS

Laryngostroboscopic examinations resulted in 542 recordings that could be used for further analysis. As seen in Table 1, only a small minority of the subjects was able to accomplish the whole set of tasks. Moreover, a small number of recordings could not be used because one of the observers did not give a score.

During the accomplishment of tasks the subjects phonated at freely chosen pitches. Phonations in a falsetto register were avoided, whenever possible. Table 2 gives the averaged fundamental frequencies and standard deviations for both women and men. The frequencies are in close range for each separate frequency level, with the exception of the phonation with low pitch at normal intensity level, which has a considerably higher frequency. Absolute values for sound pressure levels were not obtained due to the automatic gain control of the recording equipment.

Inter-observer reliability for rating the percentage of glottal closure was determined using Cronbach's α, which ranges from 0, indicating no agreement, to 1, indicating complete agreement. A high level of agreement of α=0.93 was established for rating glottal closure. The mean inter-observer correlation was 0.82. Intra-observer reliabilities were established with correlation coefficients. The correlation coefficients varied among the three observers from 0.63 to 0.98 ($p<0.001$).

Intensity level	Frequency level		
	Low	Normal	High
Men			
Soft	68.6 36.55 (19)	83.7 15.38 (27)	72.7 25.98 (27)
Normal	92.2 9.65 (36)	90.2 19.34 (41)	89.7 19.94 (38)
Loud	99.3 2.12 (21)	99.2 2.38 (29)	96.5 8.30 (29)
Women			
Soft	67.0 20.00 (10)	67.5 24.44 (17)	45.1 23.92 (14)
Normal	88.4 12.85 (19)	83.6 13.22 (80)	70.6 20.97 (77)
Loud	96.5 5.09 (18)	95.2 5.35 (29)	88.1 14.89 (28)

Table 1. Averaged percentage of glottal closure
during phonatory tasks, and standard deviations for men and women.
( ): number of observations.

Intensity level	Frequency level		
	Low	Normal	High
Men			
Soft	127.1 (13.68)	171.9 (17.78)	249.8 (37.70)
Normal	121.9 (18.21)	175.1 (27.34)	263.1 (55.06)
Loud	128.4 (19.39)	173.9 (21.75)	254.8 (67.52)
Women			
Soft	166.8 (33.53)	251.1 (32.68)	385.4 (52.81)

Table 2. Averaged frequencies in Hz produced during phonatory tasks for men and women.
( ): standard deviations.

Table 1 gives the averaged percentage of closure for each intensity and frequency, separately for men and women. The averaged percentage of closure varies between 68.6 to 99.3% in the men and between 45.1 and 96.5% in the women. The lowest values are found for phonation with high pitch and low intensity, and the highest values for phonation with low pitch and high intensity. There is a considerable difference in the magnitude of the standard

	Glottal closure					
	Women			Men		
	df	F	p	df	F	p
Intensity	2,272	46.77	<0.001*	2,264	31.00	<0.001*
Frequency	2,272	21.67	<0.001*	2,264	1.36	0.259

Table 3. Analysis of variance summary table with effects of intensity and frequency on percentage of glottal closure. *$p<0.05$. df = degrees of freedom.

deviation. The highest standard deviations are found for the low intensity levels, while there are relatively small standard deviations for the high intensity levels. This means that a large variety in percentage of glottal closure can be found at a low intensity level, whereas this variation is limited at high intensities.

To establish differences in glottal closure between men and women, as well as to determine the effects of frequency and intensity level, three way analysis of variance was performed. A significant difference was found between men and women ($F(1,540)=49.16$, $p<0.001$), with men showing a higher percentage of closure. Because significant differences were established between men and women and both intensity and frequency level ($F(2,539)=3.96$, $p=0.02$, and $F(2,539)=7.82$, $p<0.01$, respectively), the effects of these factors were separately analyzed for men and women with two way analysis of variance. Table 3 gives the analysis of variance summary table. No significant interaction between intensity and frequency level was found. In women a significant effect of both intensity ($F(2,272)=46.77$, $p<0.001$) and frequency ($F(2,272)=21.67$, $p<0.001$) was established, whereas in men only a significant effect of intensity ($F(2,264)=31.00$, $p<0.001$) was found. Glottal closure improves in both sexes with increasing intensity, and, specifically in women, with decreasing frequency.

DISCUSSION

Though incomplete glottal closure is described in a number of articles, specific information on the completeness of closure is sparse. Södersten et al.

give quantitative descriptions of glottal closure on a discrete scale (1). Ratings of glottal closure using a percentage supply data that can be used to observe small changes in closure and to analyze the influence of variables such as vocal pitch and intensity. However, before introducing a new method to describe glottal closure the applicability and accuracy should be determined. In this study a high inter- and intra-observer reliability was established, giving evidence of the clinical practicability.

A feature of glottal closure not expressed in rating with a percentage is the location of the incomplete part. Most incomplete closures are located posteriorly, especially in women (1), whereas in men sometimes an unclosed glottis anteriorly can be seen (2). However, this absence of location of the glottal gap presents no obstacle in using the information on percentage in analyzing relationships with physiological variables and quantifying the robustness of the larynx.

Essential for an adequate assessment of glottal closure is the equipment used in visualizing the larynx. A 90° rigid endoscope with a powerful light source provides an undistorted and detailed laryngeal image. During phonation, stroboscopy shows vocal fold functioning and closure can be closely observed. The image obtained should be recorded in order to make it possible to analyze the images in more detail --if necessary in slow motion--afterwards.

The percentage of glottal closure with standard deviations indicate that, especially in women at lower intensities, the majority of subjects do not have complete closure during phonation. This is in agreement with other studies giving information on the relative number of complete glottal closures among subjects (1,3,4).

The results of this study demonstrate better closure in men compared to women, as well as the specific influence of both pitch and vocal intensity on glottal closure. Better closure in men is in agreement with the results of Södersten et al. (1), who also reported a positive effect of vocal intensity on closure in both sexes, although in women this effect was not significant with a p-value (p=0.0106) slightly above their chosen level of significance (p<0.01). Although not significant, in women a negative effect of increasing pitch on glottal closure was also observed by Södersten et al., which confirms the finding of the present study. The smaller number of subjects used in the study of Södersten et al. is presumably responsible for the fact that observed influences on glottal closure are not significant in all cases.

Because variations in vocal intensity and, to a lesser extent only in women, variations in pitch have a significant influence on glottal closure, the functioning larynx should not be evaluated at only one pitch and intensity level. Variation in pitch and intensity exemplifies the effect of vocal fold physiology on glottal closure. Extremes in phonatory conditions, that is loud phonation with normal or low pitch, in contrast to quiet phonation with high pitch, however, outline the anatomical restrictions of the voice source.

In clinical practice an almost complete glottal closure of at least 90% has to be observed during loud phonation in women, whereas this closure should be complete in men. If these specifications are not reached, it raises the presence of a less robust larynx, which is more susceptible to vocal complaints (5,6). Diagnosing a less robust larynx may therefore not only have consequences for voice entrance examinations to studies that require an optimal vocal apparatus of the candidate, regarding the intensive use of the voice (schools for singers, actors and speech therapists), but it can also lead to specific advice regarding choice of a profession. To distinguish a less robust larynx from a normal larynx, glottal closure should be judged at several intensities, from quiet to loud.

Potentially beneficial treatments for the less robust larynx are limited. This should be kept in mind while giving advise regarding the choice of profession. To promote a responsible voice use and to prevent secondary vocal fold abnormality, speech therapy is advised. During a limited number of sessions the patient can explore the limitations of his or her voice and learn to optimize voice possibilities. With the knowledge of the voice possibilities prolonged vocal hygiene is pursued, which helps to minimize voice strain and thereby the risk of vocal fold damage.

References

1. Södersten M, Lindestad P.-A. Glottal closure and perceived breathiness during phonation in normally speaking subjects. J Speech Hear Res 1990;33:601-11.
2. Sulter A, Schutte H, Miller D. Standardized Laryngeal Videostroboscopic Rating: Differences

between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency and age. J Voice 1996;10:175-89.
3. Biever D, Bless D. Vibratory characteristics of the vocal folds in young adult and geriatric women. J Voice 1989;3:120-31
4. Linville S, Skarin B, Fornatto E. The interrelationship of measures related to vocal function, speech rate, and laryngeal appearance in elderly women. J Speech Hear Res 1989;32:323-30.
5. Morrison M, Rammage L, Belisle G, Pullan C, Nichol H. Muscular tension dysphonia. J Otolaryngol 1983;12:302-6.
6. Dikkers F, Nikkels P. Benign lesions of the vocal folds - Histopathology and phonotrauma. Ann Otol Rhinol Laryngol 1995;104:698-703.

4

Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency and age

Arend M. Sulter, Hero P. Wit

Department of Otorhinolaryngology, University Hospital Groningen, Groningen, the Netherlands

Accepted for publication by the *Journal of the Acoustical Society of America*.

INTRODUCTION

Voice evaluation is an important topic for assessing a person's vocal abilities. Evaluation implies a frame of reference. Data bases play an important role in this respect. This study is the result of an extensive research project, the aim of which is to supply information on voice characteristics of a large group of subjects with healthy voices. Related studies have already established normative data for phonetograms (1), laryngostroboscopy (2), and glottal closure (3). In this study, normative data on glottal flow characteristics are established for the evaluation of voice production.

Vocal sound is based on the generation of pressure differences in the larynx. These pressure differences are the result of vibratory actions of the vocal folds on the expiratory air flow from the lungs. The cyclical opening and closing of the vocal folds is influenced by physiological and aerodynamic factors, such as subglottal pressure, vocal fold tension, and elastic and

Bernoulli forces (4-6). Characterization of the glottal volume velocity waveform therefore provides useful information on phonatory function (7-9). Measurements of glottal flow can also provide benchmark data for voice source models.

Examination of glottal function during voice production has been problematical due to the relative inaccessibility of the larynx in the human body for investigational procedures. However, since the introduction of the circumferentially vented pneumotachograph (10), measurements on glottal flow can be easily --as a non-invasive procedure-- performed, reflected in numerous publications. [1] Glottal waveform characteristics have been described in normal adults (11-15) as well as in children (16,17), in aged adults (18), and in patients (19-25). Furthermore, correlations with intensity (11, 26-28), voice type (8,24), and singing technique (29,30) were investigated.

Differences in glottal waveform characteristics between singers and untrained subjects have also been studied. Trained singers are accustomed to exploiting a fuller intensity and pitch range, which requires laryngeal adjustments (31) with specific respiratory (32) and vocal function (33). Sundberg and Gauffin (15) and Sundberg and Rothenberg (34) measured higher peak flows in singers, presumably reflecting differences in glottal adductory forces. However, both these studies were performed with a limited number of subjects.

In the present study glottal waveform characteristics of a large group of subjects with and without vocal training were determined to create a data base with normative values. Normative data on untrained, i.e., "normal" subjects might function as a frame of reference for future investigations. Subjects with vocal training were included to offer information about possible "good" glottal characteristics, assuming that glottal waveform characteristics can range from poor to excellent.

To determine potential influences of variables on waveform characteristics, effects of the factors gender, vocal training, sound intensity level, pitch and age were also analyzed.

[1]Most of the direct measurements of the voice source rely on invasive procedures, thereby compromising the integrity of the body. The oral flow is acquired with a non-invasive procedure, which makes it easy to perform. However, the determination of glottal flow is a rather complex matter, regarding the inverse filtering procedure.

METHODS

Subjects

A total of 224 Dutch untrained and trained subjects of both genders, categorized accordingly into 4 groups, were investigated. The untrained subjects in the first two groups were recruited from groups of students and volunteers without complaints of or a history of vocal pathology. Group 1: untrained females, n=92, age 17-44, mean 20.3, median 19, standard deviation (SD) 7.37 years. Group 2: untrained males, n=47, age 17-35, mean 25.0, median 25, SD 4.68 years. Eighteen of the female, and 16 of the male subjects were smokers. [2]

Groups 3 and 4 consisted of amateur singers with a minimum of two years of vocal training. Vocal training consisted either of singing in a choir that organized rehearsals at least once a week, or receiving individual singing lessons with a similar frequency. All the choirs had a professional conductor and used auditions to admit new members. Although a minimum of 2 years of organized singing was used as a selection criterion for the trained group (cf. Teachey et al. [33]), about 60% of the trained subjects had a considerably longer history of singing in a choir (> 5 years). Group 3: trained females, n=42, age 18-59, mean 35.1, median 34, SD 11.86 years. Group 4: trained males, n=43, age 21-75, mean 47.5 median 49, SD 18.52 years. Five of the female, and 11 of the male trained subjects were smokers. Before the actual investigation took place, all subjects were examined laryngostroboscopically to exclude vocal fold pathology. Because we depended on volunteers it was not practicable to match the gender groups according to age.

Speech material

Each subject was asked to perform a set of phonatory tasks. The tasks comprised a word, a sentence and a CVC sequence, which were produced at three sound intensity levels: soft, comfortable (hereafter referred to as normal), and loud. The intensity levels were chosen by the subject with the investigator's approval, excluding whispering and shouting. Both the word /stagiære/ (trainee) and the sentence /hou eens op te blæren/ (stop bawling) contained the stressed vowel /æ/, which had to be slightly elongated. The vowel /æ/ was

---

[2]Smokers were included in the research groups to use a normal representation of the Dutch population. The percentage of smokers in the research groups was comparable to the percentage of smokers in the normal population. As in the non-smokers, vocal pathology in the group of smokers was excluded by a close videolaryngostroboscopic examination. The number of male and female smokers did not significantly differ between the untrained and trained groups.

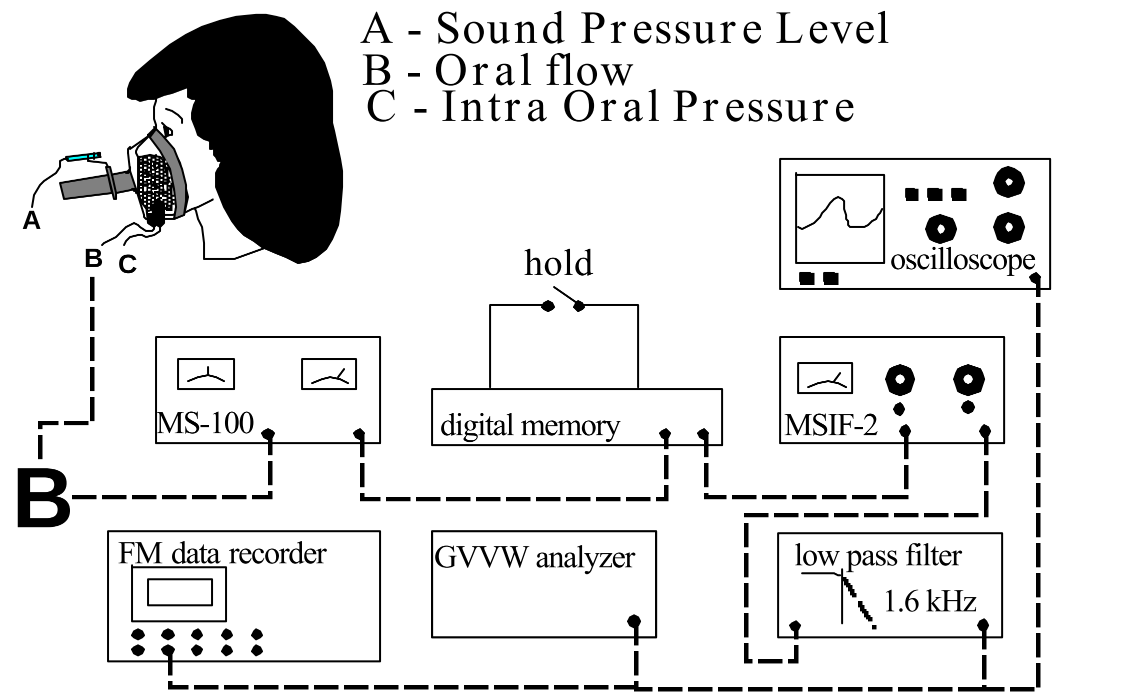


Figure 1. Experimental setup. Oral flow (B) is measured with a Rothenberg mask and the MS-100 unit. The output is stored in a digital memory. The content is read out repetitively and manually filtered with the MSIF-2 unit. After low pass filtering (1.6 kHz), the inverse filtered signal is monitored on the oscilloscope and characteristics of the flow can be observed on the GVVW analyzer. After determining a proper setting of the filters, the glottal volume velocity waveforms are recorded.

chosen for its high first formant and the separation in the frequency range from the second formant (7), as well as the "neutral" shape of the vocal tract during the production of this vowel. The high first formant diminishes interaction with the fundamental frequency, which facilitates the inverse filtering procedure. The "neutral" shape of the vocal tract minimizes the voice source - vocal tract interaction (10,14,35). Both allowing the subject to use their comfortable levels of intensity, and the use of a word and a sentence as phonation tasks, presumably resulted in a relatively natural voice production.

Immediately following the utterance of the word and the sentence the subjects phonated the CVC sequence /bæpbæpbæp/ at a rate of 4 syllables per second in each of the intensity conditions. The CVC utterance was produced at the same intensity level as the preceding word or sentence. A correct performance was checked by monitoring sound pressure levels. Intra oral pressure was taken as a representative measure for subglottal pressure (36-39).

Recorded signals

Glottal flow acquisition. Glottal volume velocity waveforms (GVVW) were acquired using a single layer circumferentially vented pneumotachograph with matching pressure transducers (Glottal Enterprises), in combination with the Glottal Enterprises MS-100 and the inverse filtering MSIF-2 units, a 14-bit digital memory with a sampling frequency of 40 kHz (Cutec CD-425), and a custom-built analog GVVW-analyzer to show flow-based parameter values (see Figure 1). For specifications of the mask, as well as the theory behind the specific approach of inverse filtering, the reader is referred to Rothenberg (10,37). By activating a hold circuitry with a foot switch, the investigator could store a selected part of the oral flow signal, coming from the MS-100 unit, in the digital memory. The memory content of 400 ms was read out repetitively to the inverse filtering unit MSIF-2.

Sound pressure level. The audio signal was registered with a Sennheiser Back-Elektret-Kondensator-Mikrofone MKE 2, mounted on the mask at a distance of 7 cm from the mouth (see figure 1). The Sound Pressure Level (SPL) was derived from the audio signal with an integration time of 100 ms. A dB(A) filter was used to exclude low frequency background noise from contributing to the SPL. The placement of the mask between mouth and microphone was experimentally determined to result in an attenuation of the SPL with 5 dB.[3]

Intra oral pressure measurement. To measure the intra oral pressure (IOP), a 2 cm removable piece of a Charière 8 silicone suction catheter was connected with an adapting tube to a differential pressure transducer (Glottal Enterprises) with a flat frequency response up to about 30 Hz. The other end of the tube was placed in the oral cavity. The position prevented the tube from being filled with saliva during the experiment, while the plasticity of the tube material allowed unhampered conditions for the accomplishment of the speech tasks.

Inverse filtering

To compensate for the resonances of the vocal tract, the digitally stored oral flow signal of 400 ms was manually inverse filtered by monitoring on the oscilloscope (Hameg Digital Storage Scope HM 208) the result of adjusting the two inverse filters of the MSIF-2 unit.[4] The goal of adjusting the filters was to

---

[3]A small experiment was conducted to obtain information on the attenuation of the mask. The mask with the microphone was placed in the experimental design described in the Methods section. The mask was then removed, while keeping the distance from microphone to artificial voice constant. The difference in SPL between both measurements was 5 dB.

[4]The use of Rothenberg's method for inverse filtering calls for a critical attitude (67). In the process of inverse filtering, well described criteria were used to adjust the two inverse filters with centre frequencies and bandwidths, resulting

arrive at a maximally flat portion of that part of the GVVW, which represents the closed phase of the glottal cycle (40). In a number of cases in male subjects the most optimal setting of filters resulted in a hump of the waveform at the beginning of the most closed phase (cf. 26,35,41). To remove high frequency energy, the derived glottal flow signal was low pass filtered (Frequency Devices 8 pole Bessel 902 LPF) with a cut-off frequency of 1.6 kHz, according to the resonance characteristics of the mask (42).[5]

Registration of signals

The glottal flow signal, the sound pressure level and the intra oral pressure signal were registered on VHS tapes with an instrumentation recorder (TEAC XR-5 10 cassette data recorder) at a speed of 38.1 cm/s, offering an effective frequency range from DC to 10 kHz.

Calibration

Flow and sound pressure level Before each measurement session the equipment was calibrated for flow and sound pressure levels. These calibration signals were recorded along with speech signals for each subject. The flow mask with its differential pressure transducer was calibrated at three air flow rates, namely, 0, 400 and 800 ml/s, by placing the mask with a tight seal against an artificial head that had a laminar flow connection with a central air supply. The exact flow level could be adjusted by means of a Brooks 2-tube sho-rate flow meter.

The sound pressure level was calibrated at 70, 75 and 80 dB by placing the mask on a mould, which incorporated the B&K Artificial Voice Type 4219. The artificial voice was driven by the B&K Beat Frequency Oscillator Type 1022 at a frequency of 150 Hz.

Pressure. The transducer for intra oral pressure measurements was calibrated with a water manometer at the levels 0, 10, 20, 30 and 40 cmH_2O before the first subject was investigated. A drifting of the transducer characteristics was checked from time to time but never showed any significant deviation from the original calibration curve.

---

in high intra-researcher correlation.

[5]The acoustic properties of the mask used in this study were investigated with the equipment described in the pertinent section. The obtained resonance characteristics were in agreement with those given by Hertegård and Gauffin (42).

Data acquisition

The subject was asked to push the mask firmly against the face, expire, and to explore any leakage of air other than through the mesh wire screen incorporated in the mask. The same mask was used for all subjects and cleaned between recording sessions. In a number of females the back of the nose was too small to fit properly in the mask. In those cases that part of the mask was filled with a mouldable silicone based impression material (Optosil P plus; Bayer Dental). During the experimental tasks, an incorrect position of the mask, resulting in unintended air-leakage, could be monitored with a zero-level indicator and a connected flashing red light on the analog GVVW-analyzer. In case of a leakage, the DC-component of the GVVW is reduced to zero and hence the remaining AC-component indicates an erroneous measuring condition.

Before the actual registration, the phonation tasks were practised with the guidance of the investigator. Of each task, a first recording to register speech signals started at the beginning of the utterance of the subject and ended after activating the hold-circuitry. After the beginning of the /æ/ vowel the hold switch was activated and a 400 ms midvocalic oral flow signal was stored in the digital memory for inverse filtering and subsequent determination of the glottal flow signal. The stored signal was checked for a steady state appearance by comparing the levels of the amplitudes of the signal on the oscilloscope and by listening to the stored signal over headphones to verify the vowel quality. To remove the formant ripple the filters were adjusted manually, while checking the GVVW on the oscilloscope. The final centre frequencies and relative bandwidths were written down in order to trace questionable filter settings. After the completion of the filtering procedure a second recording was made onto tape of the inverse filtered signal.

The IOP signals produced during the phonation of the CVC sequence were registered, while monitoring the excursions of a VU-meter connected to the IOP-transducer to check for a correct performance of the task and a proper function of the equipment.

All examinations were performed by the same investigator.

Signal processing and data analysis

Digitization. The recorded signals were digitized with a 12 bit successive approximation converter (MetraByte DASH 16). A sampling frequency of 500 Hz was used to convert the calibration and the speech signals. The inverse filtered signals were digitized with a sampling frequency of 10 kHz. All the signals were digitized simultaneously and then demultiplexed. The inverse filtered signal was stored in subfiles of 2048 samples, which corresponds to approximately 0.2 seconds.

Analysis. A parameter extraction program was written in a fourth generation signal analysis language (ASYST, MacMillan Software Company) to analyze sound pressure level, intra oral pressure level and glottal volume velocity waveform parameters. Individual calibration files were used to quantify signals.

GVVW parameters. For each utterance parameter extraction was performed on a representative subfile of 2048 samples. Depending on F_0, about 20 to 60 cycles were analyzed with fundamental frequencies ranging from 100 Hz to 300 Hz, respectively.

A peak-picking algorithm was used to identify the fundamental period T (see figure 2) and derive the fundamental frequency. Maxima of the GVVW signal and the second derivative of this signal were used to determine maximum flow (A in figure 3), and moments of opening (B in figure 3) and closure (C in figure 3) of the glottis, respectively, within a single fundamental period. After labelling events A, B and C, closed time (t_1), opening time (t_2) and closing time (t_3) (see figure 2) were calculated. Closed quotient (t_1/T), closing quotient (t_3/T) and speed quotient (t_2/t_3) could be determined

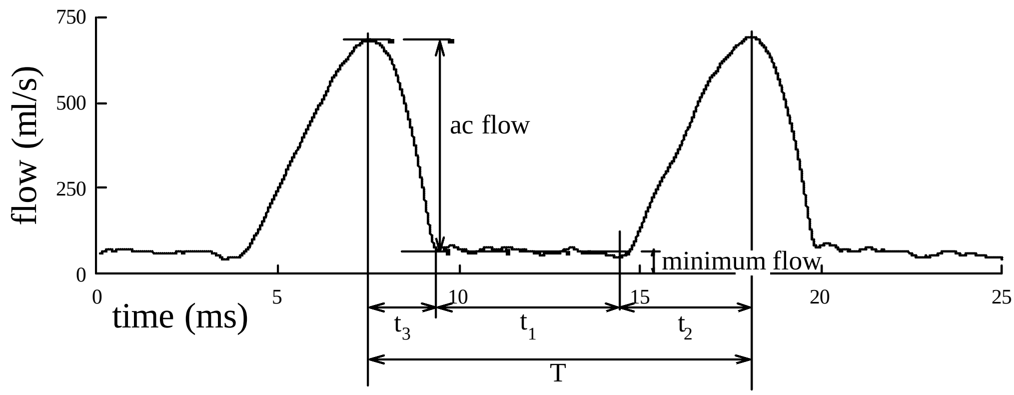


Figure 2. Glottal volume velocity waveform parameters and markers.
T = fundamental period, t_1 = closed phase, t_2 = opening phase and t_3 = closing phase.

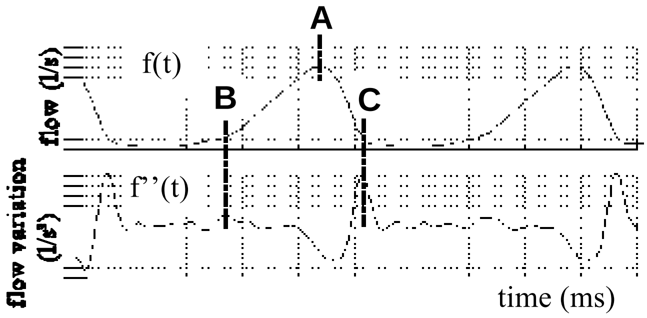


Figure 3. Glottal volume velocity waveform and its second derivative. Markers indicate maximum flow (A), moment of opening (B), and moment of closing (C).

consequently.[6] Minimum flow was calculated by averaging flow over the most closed period (t_1). Subtracting minimum flow from maximum flow yielded ac flow. Average flow was calculated by taking the mean of the glottal flow. Maximum flow declination rate (MFDR) was calculated by determining the minimum of the first derivative of the glottal waveform.[7]

Finally, two parameters related to vocal fold function were calculated. Vocal efficiency was determined by calculating the ratio between sound power and the product of air pressure times average flow (43). A glottal resistance measure was determined by calculating the ratio of air pressure to average flow (38).

Intra oral pressure adjustment. After analyzing the CVC sequence and determining intra oral pressure (IOP) and SPL, IOP was standardized for the SPL produced during the utterance of the preceding word and sentence. Regression analysis of individual SPL-IOP functions showed a linear relation with a high correlation coefficient (r>0.98) between SPL and the logarithmically transformed IOP values. Regression coefficients were used to predict the IOP pertaining to the utterances.

[6]The method of inverse filtering presented in this study supplies waveforms with minimal variation of flow during the closed portion of the glottal cycle, as this minimal variation is the main criterion for accepting the filter settings. With the subsequent low pass filtering a waveform results, which generally presents no problem for the algorithms in detecting representative maxima in the second derivative. Soft phonations with very limited modulation sometimes give problems and lead to rejection of some waveforms for further analysis.

[7]Representing the maximum in the change of flow, MFDR was regarded as a flow-based parameter.

Statistical analysis

Analysis of variance (ANOVA) and Multivariate analysis of covariance (MANCOVA) of the statistical package SPSS (SPSS Inc.) was used to investigate differences among groups (44). Minimum flow, ac flow, average flow, MFDR, closed quotient, closing quotient, speed quotient, glottal resistance and vocal efficiency were regarded as dependent variables. Gender, vocal training, and intensity condition were introduced as factors. SPL, fundamental frequency, IOP, and age served as covariants. In case of significant interaction between factors, a separate MANCOVA was performed at each factor level. To determine differences among factor levels of the intensity condition, one-way analysis of variance with post hoc Least Square Difference (LSD) tests was performed. Because of the many tests performed, a conservative probability level $\alpha=0.01$ was used with respect to the Bonferroni inequality.

	male subject				female subject			
	soft	normal	loud	eta	soft	normal	loud	eta
minimum flow (ml/s)	198 (2.2)	69 (3.1)	143 (14.5)	0.99	112 (6.7)	84 (4.6)	81 (2.3)	0.95
ac flow (ml/s)	245 (2.0)	497 (5.6)	805 (13.0)	1.00	170 (1.4)	209 (3.6)	384 (17.4)	0.99
average flow (ml/s)	263 (43.7)	153 (30.9)	329 (8.6)	0.93	172 (5.5)	139 (9.0)	170 (17.8)	0.80
MFDR (l/s^2)	210 (16.9)	564 (17.8)	1863 (131.1)	0.99	218 (6.6)	495 (34.7)	1060 (64.1)	0.99
closed quotient (%)	42.3 (1.16)	48.2 (1.62)	58.7 (2.63)	0.97	32.1 (0.57)	45.7 (4.35)	54.5 (0.53)	0.97
closing quotient (%)	23.6 (0.84)	18.2 (0.63)	18.5 (0.71)	0.96	28.5 (0.53)	20.4 (0.84)	24.1 (0.57)	0.98
speed quotient	1.45 (0.096)	1.86 (0.143)	1.307 (0.100)	0.91	1.39 (0.037)	1.67 (0.268)	0.89 (0.042)	0.91

Table 1. Variance in glottal flow parameters for one male and one female speaker after ten times manually inverse filtering the same oral waveform. Mean and standard deviations (between brackets) are given. Eta values represent the strength of association between the averaged value of the parameter and condition (maximum value = 1, minimum = 0). MFDR = Maximum Flow Declination Rate.

RESULTS

Inverse filtering procedure

The variability of inverse filtering, which might result in arriving at different parameter values for GVVW, was determined with a small experiment. A male and female subject uttered the word /stagiære/ at the three intensity conditions. At each condition the investigator adjusted the filters from a neutral position to an optimal setting ten times. A registration was made of the filtered waveform and the procedure described above was used to analyze GVVW. Table 1 gives the mean parameter values with standard deviations. In general, very low standard deviations were found for all parameters, which indicates the robustness of the manually adjusted inverse filtering procedure. The eta values show the high level of association between parameter values and intensity condition.

Glottal volume velocity waveform analysis

A small number of filtered waveforms (<3%) could not be analyzed or used for further evaluation. These were mostly soft voice productions resulting in waveforms with very limited modulation, which offered insurmountable problems for the parameter extraction program. In other cases the IOP signal did not conform to the typical pattern of alternating zero - non zero pressure levels, which led to rejection of the matching GVVW for further analysis. A total of 1308 analyzed cases could be used for producing summary and inferential statistics.

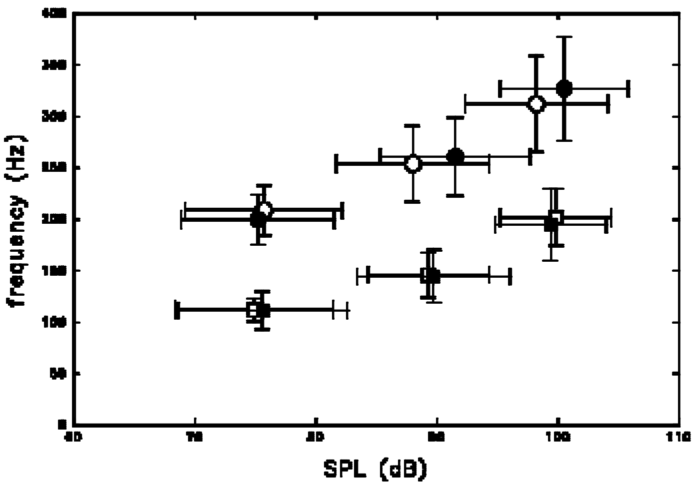


Figure 4. Relation between SPL and fundamental frequency (F_0) for soft, normal and loud intensity. Mean (○=untrained women; ●=trained women; □=untrained men; ■=trained men) and standard deviations (⊣) are given.

Summarized data

All data on glottal waveform parameters and other analyzed variables from both utterances were averaged according to gender, vocal training and intensity condition. Mean values with standard deviations are given in Table 2. This table offers normative data on vocal function and can be used to compare with other published data on

GVVW characteristics.

Averaged for intensity condition, the sound pressure levels were comparable among groups. Figure 4 presents the relation between fundamental frequency and SPL. Soft voice was produced at about 75 dB, normal voice at about 90 dB, and loud voice at about 100 dB. The increase in intensity from soft to normal voice was larger than the increase from normal to loud (15 dB versus 10 dB, respectively). A three-way analysis of variance with gender, vocal training and intensity condition as factors showed a statistically significant influence of training on SPL ($F(1,1297) = 8.47$, $p=0.004$; trained subjects produced louder phonations than the untrained ones), as well as significant differences in SPL among condition levels ($F(2,1296) = 1711.55$, $p<0.001$). The effect of gender was not significant ($F(1,1297) = 0.19$, $p=0.661$). Except for the loud intensity condition, the measured fundamental frequencies were within the ranges observed in normal Dutch speech (45). The loudest condition demonstrated frequencies within the area of chest register voice production (1). Three-way

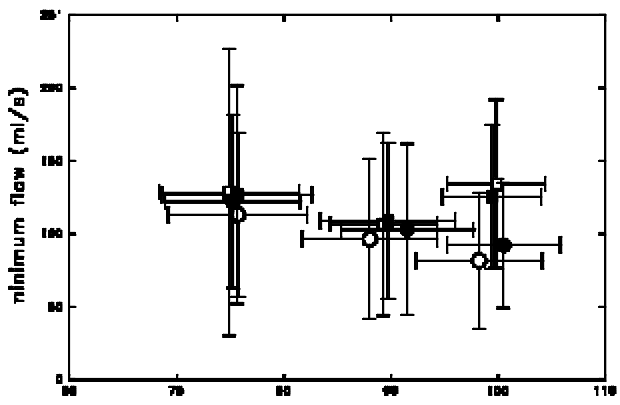
	female						male					
	untrained			trained			untrained			trained		
	soft	normal	loud	soft	normal	loud	soft	normal	loud	soft	normal	loud
vocal intensity (dB) [a]	75.7	88.0	98.2	75.2	91.5	100.5	74.9	89.3	99.8	75.6	89.7	99.4
	(6.51)	(6.29)	(5.93)	(6.34)	(6.18)	(5.27)	(6.51)	(4.99)	(4.06)	(7.05)	(6.28)	(4.61)
frequency (Hz)	208.9	253.5	312.3	200.0	261.3	326.9	112.2	145.5	201.7	111.6	144.5	194.6
	(24.30)	(36.96)	(46.42)	(24.10)	(38.11)	(50.74)	(10.86)	(21.87)	(27.55)	(18.35)	(25.69)	(34.91)
minimum flow (l/s)	0.11	0.10	0.08	0.12	0.10	0.09	0.13	0.11	0.13	0.13	0.11	0.13
	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.04)	(0.10)	(0.06)	(0.06)	(0.07)	(0.05)	(0.05)
ac flow (l/s)	0.17	0.26	0.34	0.17	0.30	0.37	0.33	0.57	0.75	0.40	0.60	0.76
	(0.06)	(0.09)	(0.10)	(0.06)	(0.10)	(0.10)	(0.10)	(0.15)	(0.16)	(0.12)	(0.18)	(0.16)
average flow (l/s)	0.16	0.16	0.17	0.17	0.18	0.18	0.21	0.23	0.32	0.23	0.24	0.31
	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)	(0.12)	(0.08)	(0.09)	(0.09)	(0.09)	(0.09)
maximum flow declination rate (l/s^2)	249.7	504.0	782.8	265.7	677.2	966.8	393.9	1026.2	1689.1	429.3	1043.7	1641.5
	(112.4)	(220.1)	(243.0)	(108.4)	(270.5)	(283.9)	(192.7)	(372.0)	(397.4)	(211.0)	(431.9)	(374.2)
closed quotient (%)	35.3	44.5	47.5	36.0	45.8	48.6	42.4	50.8	49.3	41.0	49.4	50.1
	(9.4)	(11.6)	(11.5)	(10.5)	(10.9)	(10.6)	(12.2)	(9.1)	(8.3)	(11.1)	(8.5)	(8.0)
closing quotient (%)	25.8	23.7	24.3	26.5	24.8	26.6	22.4	19.7	21.0	22.4	19.0	19.6
	(4.2)	(3.2)	(3.7)	(4.4)	(3.7)	(5.1)	(5.1)	(3.5)	(3.5)	(5.0)	(2.7)	(3.1)
speed quotient	1.55	1.36	1.19	1.45	1.20	0.97	1.63	1.52	1.46	1.68	1.67	1.58
	(0.38)	(0.48)	(0.49)	(0.33)	(0.39)	(0.44)	(0.55)	(0.35)	(0.42)	(0.36)	(0.40)	(0.41)
pressure (cm H_2O)	6.0	11.2	18.3	5.5	13.2	21.0	5.6	11.8	20.5	5.9	12.5	20.9
	(2.2)	(4.1)	(6.9)	(2.3)	(5.3)	(6.4)	(2.4)	(3.7)	(5.9)	(2.7)	(5.3)	(6.3)
glottal resistance (cm H_2O/l/s)	45.6	81.8	123.9	37.4	91.0	130.3	36.9	56.8	70.0	30.6	64.4	80.1
	(31.0)	(47.6)	(67.2)	(23.2)	(75.1)	(54.6)	(28.2)	(26.4)	(26.8)	(20.0)	(88.1)	(65.4)

Table 2. Summary statistics of variables and glottal waveform parameters. Mean and standard deviation (between brackets) are given.

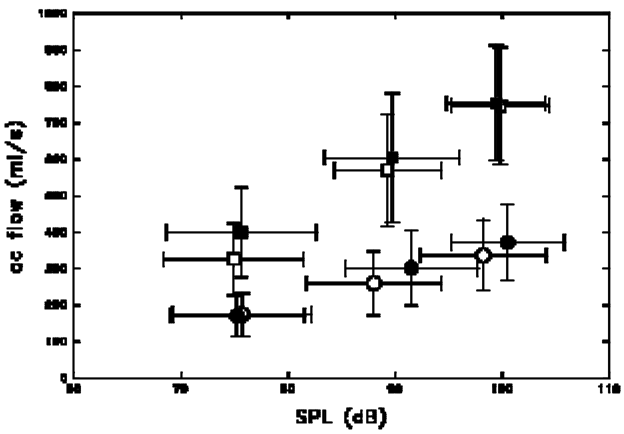
[a]Measured at 7 cm from the mouth

[b]Vocal efficiency values have been multiplied by 10^5.

analysis of variance of fundamental frequency showed a statistically significant interaction between the factors gender, vocal training and intensity condition ($F(2,1296) = 5.72$, $p=0.003$). Therefore, a separate two-way analysis was performed for male and female subjects. The fundamental frequency of the male subjects was significantly influenced by intensity condition, with louder conditions showing higher frequencies ($F(2,524) = 806.55$, $p<0.001$). No influence of training was observed ($F(1,525) = 0.24$, $p=0.626$). In the female subjects, a statistically significant interaction between training and intensity condition was observed ($F(2,769) = 6.21$, $p=0.002$). The fundamental frequency of both the trained and untrained female subjects was significantly influenced by intensity condition, louder conditions giving higher frequencies ($F(2,261) = 237.64$, $p<0.0001$, and $F(2,550) = 349.16$, $p<0.0001$, respectively). The trained females phonated with a lower fundamental frequency during the tasks with soft voice than untrained ones ($F(1,285) = 10.78$, $p=0.0012$).



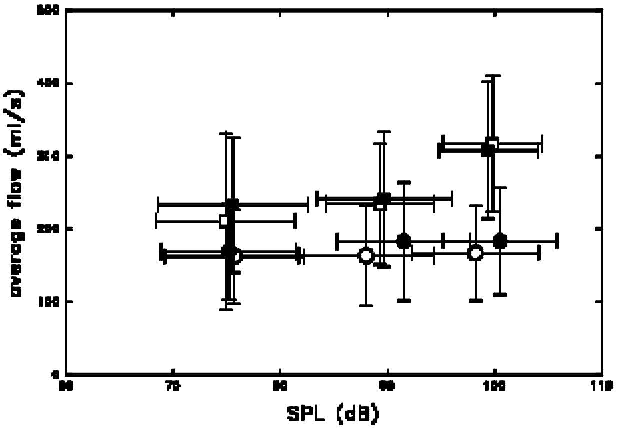
5a



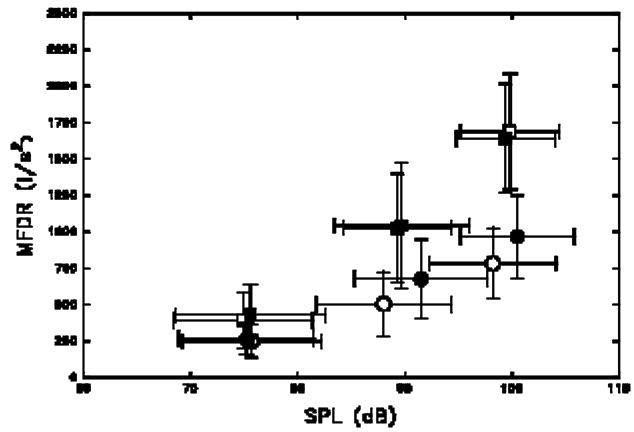
5b

To make the relation between the GVVW characteristics and the sound intensity more comprehensible, as well as to show differences between gender and vocal training, plots were constructed for each GVVW parameter. Mean values are given for the four speaker groups. Female groups and male groups are represented by circles and squares, respectively. Untrained groups have unfilled figures and the figures of trained groups are filled. The whiskers indicate the standard deviations. All presented parameters are plotted against SPL.

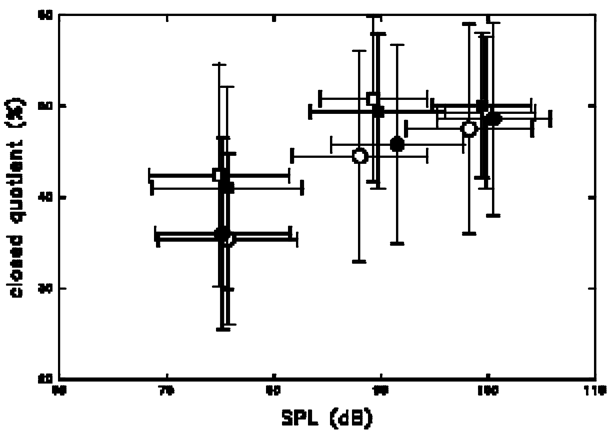
Figure 5a shows data for minimum flow, plotted against SPL. All mean values for the groups lay close together for soft and normal voice, with the male groups having the largest



5c



5d



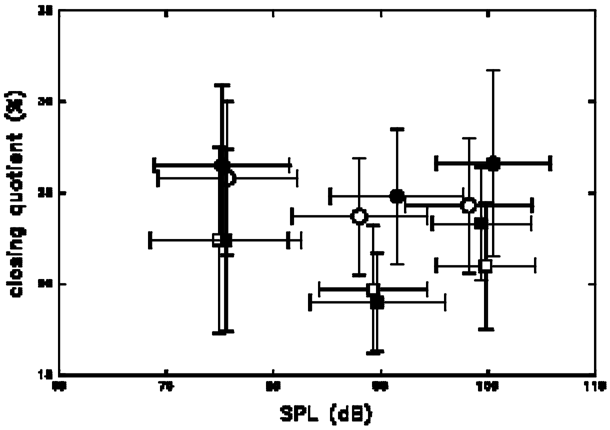
5e

standard deviations for the soft intensity condition. In loud voice, the female groups show a tendency of decreasing minimum flow with increasing intensity, whereas the male groups show an increase in minimum flow from the normal to the loud condition.

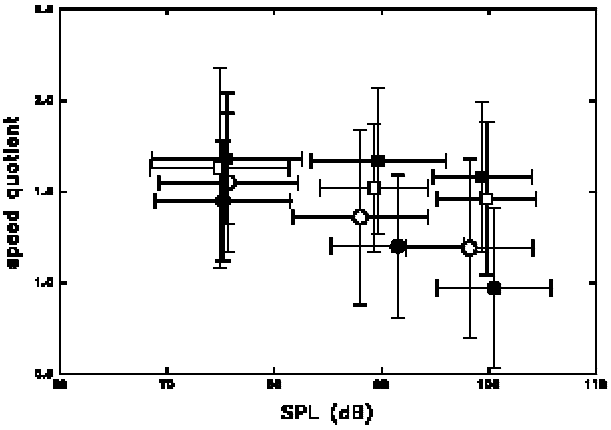
Figure 5b shows data for ac flow, plotted against SPL. Ac flow increases with intensity for all groups; the male groups, however, show the largest increments with intensity, and the ac flow in these groups is larger than in the female groups.

Figure 5c shows data for average flow, plotted against SPL. In general, the male groups have higher levels of average flow compared to the female groups, and at the loud intensity condition there is a marked increase in average flow for both the untrained and trained male group. The average flow does not show apparent changes over the intensity conditions in the female groups.

Figure 5d shows data for MFDR, plotted against SPL. In the soft intensity condition, the differences between the male and female groups are small; with increasing intensity, however, there is a distinct difference, with male subjects having the largest



5f



5g

Figure 5. Summary figures for relations between SPL and successive parameters for soft, normal and loud intensity. ○=untrained women; ●=trained women; □=untrained men; ■=trained men; ⊣ = standard deviation. a) minimum flow, b) ac flow, c) average flow, d) MFDR, e) closed quotient, f) closing quotient, g) speed quotient.

values. In all groups, the increase in mean MFDR values with SPL is apparent. The standard deviations are small at the soft intensity condition but become larger at louder conditions.

Figure 5e shows data for the closed quotient, plotted against SPL. With louder intensities, the closed quotient shows higher mean values; however, there is a difference between male and female groups. The male groups already have the highest values in the normal intensity condition, whereas the mean closed quotient further increases in the female groups from normal to loud. In the loud intensity condition, all groups have comparable closed quotients with values close to 50%.

Figure 5f shows data for the closing quotient, plotted against SPL. Compared to the male groups, the female groups have higher mean closing quotient values, which implies that closing of the vocal folds involves a larger part of the glottal cycle in the female groups. All groups have the lowest mean closing quotient values in the normal intensity condition.

Figure 5g shows data for the speed quotient, plotted against SPL. There is a trend toward a more symmetrical shape of the GVVW with louder intensity conditions for all groups; however, the female groups show this tendency more clearly, especially the trained female group with a value of 0.97 for the loud intensity condition. The large spread in the individual values in the loud intensity condition becomes apparent from the large standard deviation for the groups.

Differences between groups and influence of variables

MANCOVA was used to investigate differences between groups, as well as among intensity conditions (44). Table 3 summarizes the effects of the factors gender, vocal training and intensity condition on GVVW parameters and the derived parameters glottal resistance and vocal efficiency. Apart from the speed quotient and vocal efficiency, the factor gender shows a significant influence on all other parameters. Because of the significant interaction between gender and training, as well as between gender and intensity condition, separate MANCOVAs were performed for male and female groups. In the female groups the factor voice training has a significant influence on MFDR and closing quotient with the trained group showing higher values for both parame ters (see Table 2). In the male groups vocal training has a significan t influence

	gender x training x intensity condition df(2,1292)	training x intensity condition df(2,1292)	gender x training df(1,1292)	gender x intensity condition df(2,1292)	gender df(1,1292)	voice training		intensity condition	
						female df(1,765)	male df(1,523)	female df(2,765)	male df(2,523)
	F (p)	F (p)	F (p)	F (p)	F (p)	F (p)	F (p)	F (p)	F (p)
minimum flow	0.28 (0.758)	0.20 (0.820)	3.30 (0.070)	11.26 (0.000 *)	19.83 (0.000 *)	0.40 (0.530)	0.44 (0.506)	1.66 (0.190)	4.95 (0.007 *)
ac flow	4.03 (0.018)	1.27 (0.280)	3.25 (0.072)	76.22 (0.000 *)	339.24 (0.000 *)	0.19 (0.660)	4.04 (0.045)	5.93 (0.003 *)	9.58 (0.000 *)
average flow	1.39 (0.249)	0.79 (0.452)	0.60 (0.437)	27.48 (0.000 *)	86.28 (0.000 *)	0.35 (0.557)	1.60 (0.206)	0.903 (0.406)	13.24 (0.000 *)
MFDR	2.32 (0.099)	0.18 (0.833)	7.18 (0.007 *)	187.03 (0.000 *)	284.16 (0.000 *)	13.97 (0.000 *)	4.62 (0.032)	14.11 (0.000 *)	12.92 (0.000 *)
closed quotient	0.35 (0.703)	0.71 (0.492)	1.91 (0.167)	11.80 (0.000 *)	8.22 (0.004 *)	0.79 (0.374)	2.37 (0.124)	3.73 (0.024)	5.68 (0.004 *)
closing quotient	2.65 (0.071)	0.18 (0.834)	29.63 (0.000 *)	0.77 (0.465)	34.50 (0.000 *)	13.76 (0.000 *)	2.13 (0.145)	6.53 (0.002 *)	3.45 (0.032)
speed quotient	0.57 (0.568)	0.37 (0.693)	33.09 (0.000 *)	5.44 (0.004 *)	0.37 (0.545)	3.05 (0.081)	8.49 (0.004 *)	0.46 (0.631)	1.75 (0.174)
glottal resistance	0.80 (0.449)	1.35 (0.259)	1.17 (0.279)	27.92 (0.000 *)	32.62 (0.000 *)	0.86 (0.354)	0.15 (0.699)	2.64 (0.072)	4.01 (0.019)
vocal efficiency	1.16 (0.314)	0.45 (0.638)	1.08 (0.299)	13.26 (0.000 *)	2.74 (0.098)	0.88 (0.349)	0.48 (0.490)	22.67 (0.000 *)	17.07 (0.000 *)

Table 3. Multivariate analysis of variance summary table.
* p<0.01

only on the parameter speed quotient, trained subjects showing higher values

a

	SPL			Frequency			Pressure			Age		
	B	t-value	p	B	t-value	p	B	t-value	p	B	t-value	p
minimum flow	-1.52	-3.79	0.000*	0.08	1.34	0.180	0.94	1.83	0.067	0.47	1.99	0.047
ac flow	4.31	8.02	0.000*	-0.45	-5.46	0.000*	6.49	9.42	0.000*	0.94	2.99	0.003*
average flow	-1.13	-2.22	0.027	0.65	2.16	0.031	2.35	3.59	0.000*	0.65	2.16	0.031
MFDR	11.96	9.35	0.000*	-0.13	-0.65	0.515	13.51	8.24	0.000*	1.62	2.16	0.031
closed quotient	0.59	7.73	0.000*	-0.08	-6.45	0.000*	0.24	2.42	0.016	0.07	1.45	0.147
closing quotient	-0.30	-12.29	0.000*	0.06	17.38	0.000*	0.06	1.92	0.055	0.02	1.60	0.110
speed quotient	0.00	1.21	0.227	-0.00	-5.52	0.000*	-0.01	-3.48	0.001*	-0.00	-2.71	0.007*
glottal resistance	0.17	0.49	0.627	-0.04	-0.78	0.437	4.99	11.18	0.000*	-0.03	-0.15	0.881
vocal efficiency	7.38	8.43	0.000*	0.42	3.18	0.002*	-0.07	-0.06	0.951	-0.12	-0.24	0.812

b

	SPL			Frequency			Pressure			Age		
	B	t-value	p	B	t-value	p	B	t-value	p	B	t-value	p
minimum flow	-2.23	-3.34	0.001*	0.29	1.95	0.052	-0.36	-0.40	0.687	-0.29	-1.22	0.222
ac flow	11.86	9.25	0.000*	-0.55	-1.92	0.055	1.00	0.59	0.558	0.12	0.26	0.793
average flow	-1.75	-1.84	0.067	0.06	0.30	0.764	1.81	1.42	0.156	-0.35	-1.05	0.296
MFDR	29.07	11.75	0.000*	1.31	2.40	0.017	10.53	3.19	0.002*	-3.16	-3.62	0.000*
closed quotient	0.85	9.70	0.000*	-0.03	-1.58	0.114	-0.33	-2.81	0.005*	0.03	1.09	0.276
closing quotient	-0.46	-14.43	0.000*	0.06	9.00	0.000*	0.12	2.82	0.005*	0.01	0.56	0.574
speed quotient	0.01	3.53	0.000*	-0.01	-6.49	0.000*	0.00	0.32	0.748	-0.00	-1.63	0.104
glottal resistance	0.82	1.82	0.069	0.01	0.83	0.934	2.93	4.87	0.000*	0.03	0.17	0.867

Table 4. Regression analysis within subgroups (intensity condition and vocal training).
* $p<0.01$. a) female subjects; b) male subjects.

than untrained ones. The factor intensity condition has a significant effect on many parameters (see Table 3). In the female groups ac flow, MFDR, closing quotient and vocal efficiency, whereas in the male groups minimum flow, ac flow, average flow, MFDR, closed quotient and vocal efficiency are significantly influenced by intensity condition. Post hoc LSD tests showed significant differences among all conditions (soft, normal and loud) for the mentioned parameters in the female groups. In the male groups, there was a significant difference between soft and normal, as well as between normal and loud intensity condition for minimum flow. For average flow, significant differences were found between both soft and normal, and loud intensity condition. For closed quotient significant differences were found between both normal and loud, and soft intensity condition. Finally, significant differences among all conditions (soft, normal and loud) were found for ac flow, MFDR and vocal efficiency.

Differences in IOP values were analyzed by a three-way ANOVA with gender, vocal training and intensity condition as factors. With louder conditions the IOP increases significantly ($F(2,1296) = 960.38$, $p<0.001$), and trained subjects use higher pressures during phonation ($F(1,1297) = 10.26$, $p=0.001$). No differences were found between male and female subjects ($F(1,1297) = 2.82$, $p=0.093$).

The influence of covariants SPL, fundamental frequency, IOP and age were also investigated separately for male and female subjects in the MANCOVAs with regression analysis within subgroups (see Tables 4a and b).

Except for glottal resistance and average flow in both gender groups, and the speed quotient in the females, SPL has a significant relation with all other parameters. Minimum flow and closing quotient have a negative regression coefficient, whereas ac flow, MFDR, closed quotient, speed quotient and vocal efficiency have a positive regression coefficient.

Fundamental frequency has a significant relation with the closing quotient, speed quotient and vocal efficiency. In the females, a significant influence of fundamental frequency is also found on ac flow and closed quotient. The ac flow, closed quotient and speed quotient have a negative regression coefficient, whereas closing quotient and vocal efficiency have a positive regression coefficient.

Intra oral pressure has a significant influence on MFDR and glottal resistance in both genders, and more specifically on ac flow, average flow, and speed quotient in females, and on closed quotient and closing quotient in males. The speed quotient and closed quotient have a negative regression coefficient, whereas the other parameters mentioned have a positive regression coefficient.

Finally, age has a significant negative regression coefficient with MFDR in male subjects, and on ac flow and speed quotient in female subjects, with a positive and negative regression coefficient, respectively.

DISCUSSION

Sound Pressure Level

Voice production entails the generation of an audible signal. An important feature of the signal to make it understandable is the sound intensity level. Therefore glottal function is described in close relation with SPL. In this study clear differences in SPL were established among the intensity conditions soft, normal and loud voice. Other studies also have employed different intensity conditions to study the behaviour of the voice source; however, a direct comparison of values found in this study with values reported previously shows a clear distinction, with the present study giving higher values. Averaged SPL values measured in the present study range from about 75 dB for the soft, 90 dB for the normal, and 100 dB for the loud intensity condition, whereas Holmberg et al. (11), Perkell et al. (12) and Stathopoulos and Sapienza (27) give mean SPL values that are about 15 dB lower for the normal and loud intensity condition. An explanation for the loud intensities in the present study can be found in the placement of the microphone. In our experimental setup the mouth to microphone distance is 7 cm, whereas in most other studies a distance of about 15 cm is mentioned (e.g. 11,26,27). This difference in distance was experimentally determined to account for 7 dB. A second explanation for the higher SPL in the present study can be found in the different task contents (46-48). In the present investigation a stressed vowel in a word and sentence was used, while most investigations work with a CVC sequence (e.g., /bæpbæpbæp/). Till et al. (49), however, did not establish differences in measurement results for a sustained vowel and a CVC sequence. Despite the differences in SPL, comparisons of glottal flow characteristics obtained in this investigation were made with results from other studies, while similar qualitative (soft, normal and loud) intensity conditions were used.

Flow-based parameters

As in most other studies the following flow-based parameters were used to describe glottal functioning: minimum flow, ac flow, average flow, and maximum flow declination rate. Peak flow (cf. 11) was not introduced because it can be derived from adding minimum flow and ac flow.

Minimum flow is related to leakage of air through a continuously unclosed part of the glottis. Two different minimum flows are used in literature. Stathopoulos and Sapienza (16,27), Sapienza and Stathopoulos (17,25), Dromey et al. (26), Higgins and Saxman (18,47), and Peterson et al. (24)

determine minimum flow at the moment during the closed phase of the glottal cycle with absolute minimum flow, while Holmberg et al. (11), Hillman et al. (20,21), and Hertegård et al. (30,41) mention the flow --which is not necessarily the absolute minimum-- during the most closed phase. In our study, the minimum flow was taken as the average of the flow during the closed phase. This procedure avoids establishing extreme values for minimum flow, which might be acquired due to incorrect filter settings for removing formants in the oral flow signal. Slightly incorrect filter settings result in ripple in the glottal waveform. Also, apart from an upward movement of the vocal folds (vertical phasing, cf. 2,41) in men at low fundamental frequencies, in normal subjects no dynamic events take place at the glottal level during the most closed phase, giving a flat portion in the glottal flow. The reduction of minimum flow with increasing intensity, as observed in the present study, is probably related to the closing of the posterior part of the glottis (2,3,11). In both the untrained and trained male subjects, however, an increase in mean minimum flow is observed from normal to loud voice. Although speculative, an explanation might be that the decrease in unclosed part of the glottis in men is less evident from normal to loud voice compared to women, and that with increasing subglottal pressure from normal to loud voice, those persons with an unclosed part of the glottis show an increase in minimum flow. Another option is that with increasing subglottal pressure, subjects with a complete closure of the vocal folds during normal voice, start leaking from the normal to loud voice condition. Other explanations are based on processes concerning both inverse filtering (source-vocal tract interaction, cf. Rothenberg (50)), as well as parameter extraction (algorithmically defined labelling of the moment of closure).

Ac flow is determined by the amplitude of the vocal fold vibration in combination with the subglottal pressure (7). Mean ac flow values in this study are higher than values reported so far. Only Hertegård et al. (41) found values for normal voice resembling the values reported in this study. Reasons for this difference might be found in one of the following: the use of a vowel instead of a CVC sequence (18,47), the stress on the vowel (46,48), specific filter settings (51), cultural differences in speech production (52), or the higher --compared to other studies-- SPL values measured in the present study. SPL has a positive relation with ac flow (11,13,15,18,25). The differences in ac flow between the intensity conditions in the present study are proportional to the differences in the articles mentioned. The increase in ac flow with louder intensities correlates with the larger amplitudes of vocal fold vibration, as observed in the same subjects (2).

Mean values for average flow resemble those values found in literature (cf. 53 for a review). Previous studies showed a slightly positive relation between average flow and SPL (43,54). With increasing SPL from normal to loud voice, the average flow significantly increases in men. This increment follows from increased minimum flow and ac flow. The steady average flow of women across the conditions might be the compensatory result of a reduction of glottal leakage and an increase in vocal fold excursions from soft to loud voice, as seen in videostroboscopic ratings (2,3).

Maximum airflow declination rate represents the closing velocity of the vocal folds and serves as an indicator of the excitation of the vocal tract; it is therefore closely related to SPL (55,56). The clinical relevance of this parameter is its assumed relation with vocal fold collision forces. Relatively high levels of MFDR are associated with vocal fold pathology (17,20-22). Because absolute values of MFDR are important in this respect, the intra-individual stability of the parameter (55,57) is crucial for studies using this parameter as an indicator of susceptibility to vocal fold pathology, as is the knowledge about the frequency dependency of low pass filter settings (12,51). MFDR values established in this study are much higher than values reported hitherto. The same explanations as given for the ac flow can be given for these extreme values. However, a similar tendency of exponentially increasing MFDR values with increasing SPL was found, as in other studies (12,17,25,27).

Time-based parameters

The maxima of the second derivative of the glottal waveform, representing the most significant changes in the glottal flow, were used to define moments of closure and opening. These important events in the glottal cycle have been determined using many other criteria. Holmberg *et al.* (11) used the intersection of visually determined line tangents along the waveform to establish moments of opening and closing. Other investigators use specific impedance information from the simultaneously recorded electroglottographic signal (24,41). Also specific levels of the ac flow are used to establish moments of opening and closing. Apart from their indication of subjective temporal markers, Dromey et al. (26) use a 50% criterion level of ac flow. Stathopoulos and Sapienza (27), in contrast, use a 20% criterion level, while Higgins and Saxman (18) use a 15% level. The differences in definition of moments of opening and closing create a problem in comparing data. In practice, our definition of closing and opening uses temporal labels that occur close to those of Holmberg *et al.* (11), Hertegård *et al.* (41), Peterson *et al.* (24), as well as the subjectively determined markers of Dromey *et al.* (26). Therefore comparisons will only be made with data originating from these studies.

Instead of an open quotient, in this investigation a closed quotient is used. However, both values are supplementary since the open and closed quotient together yield 100%. The closed quotient of the glottal flow is believed to represent the time that vocal folds are maximally approximated during a glottal cycle. Apart from elastic and aerodynamical forces, the approximation is supplied by adductory muscular activity. Therefore this parameter has also been described as adduction quotient. High adduction quotients are associated with vocal strain and might cause vocal fold pathology (24). Only few of the articles referred to above provide data on the closed quotient for other than the normal intensity condition. In this condition, the mean values range from 24% in women and 40% in men (11) to 54% in a combined group (24). Our mean values of 45% for women and 50% for men fall in between these extremes. The tendency to increase closure duration with increasing intensity (11,14,26) is also visible in our data from the soft to normal intensity condition. The untrained male subjects reach an averaged maximum of the closed quotient at the normal voice condition. The increase in the closed quotient with intensity from soft to normal voice can be explained by increased adductory muscular activity and Bernoulli forces. The small differences between closed quotient values from the normal to loud intensity condition might reflect a balance between, on the one hand, an increase of adductory muscular activity, and an increase of subglottal pressure, forcing the vocal folds apart, on the other.

The closing quotient represents the time period in which maximum flow is reduced to minimum flow. The magnitude of this parameter is strongly influenced by low pass filter settings (51). Only Holmberg et al. (11) provide data for comparison. Mean closing quotients in the present study are slightly smaller at each intensity condition. Only the closing quotient for women at loud intensity condition is slightly larger (21.0% vs 19%). In our data, the closing quotient has a minimum for all groups in the normal voice condition. This minimum might indicate a preferable condition for sound production, with a high excitation of the vocal tract as a consequence of a fast closure of the vocal folds, combined with limited ac flow levels.

The speed quotient reflects the symmetry of the glottal waveform shape. A value of 1 stands for symmetry, while values larger and smaller than one correspond to a waveform with a relatively shorter and longer closing phase, respectively. The results of investigations performed hitherto show mean values larger than 1, with inconsistent relations with intensity condition (11,26). Our results show mean values >1, with a decreasing tendency with increasing SPL.

Intra oral pressure

Intra oral pressure was used as a representative measure of subglottal pressure. Since the appearance of studies validating this measurement method, data on subglottal pressure have become available in a number of publications (11,18,27,41,58). Results of the present study show mean values, higher than the values given in the pertinent articles. However, Schutte (43) measuring subglottal pressures with an esophageal balloon, presented pressure ranges that are more in agreement with the results from the present study. In all studies, a positive relation between pressure and SPL has been established.

Derived parameters

From the parameters and variables discussed above a glottal resistance and vocal efficiency were calculated. Because the intra oral pressures from the present study are high compared to results from previous publications, it was no surprise that our mean values for vocal efficiency and glottal resistance are also extreme.

The vocal efficiency is calculated as the ratio of the produced sound power to the subglottic power (see Schutte (43) page 50 for equations). Schutte (43) established vocal efficiency values ranging from 0.12 x 10^{-5} to 400 x 10^{-5} over an intensity range of 47 dB. At 70 dB (microphone to pneumotachograph outlet distance 15 cm), the efficiency varied from 1 x10^{-5} to 10 x 10^{-5}, and at 90 dB from 10 x 10^{-5} to 110 x 10^{-5}. In our data, the women surpass this highest value in the loud voice condition, the untrained subjects with a mean value of 150.3 x 10^{-5} and the trained subjects with a value of 143.4 x 10^{-5}. Holmberg et al. (11) presents values for women that are lower than the results from the present study, whereas the mean values for the male subjects are much closer. In all studies, the efficiency index is characterized by large standard deviations, indicating the large range in individual values.

A review of the literature reveals a variability in mean glottal resistance (cf. 11,27, 59,60). The present study again provides higher mean values than those reported before, especially for women. For the male subjects, the mean values of glottal resistance are close to those presented by Stathopoulos and Sapienza (27). The high values are explained by the measured high intra oral pressures. As can be observed in the other studies, the glottal resistance increases with increasing intensity, which is related to with the increment in adductory forces.

Influence of factors

Gender. There are important differences in the anatomy of the larynx between men and women (61). The male vocal folds are longer and thicker, which plays an important role in the specific physiology and acoustics of voice production in men and women (62). The observed differences in ac flow, average flow, closing quotient, and MFDR can be explained by considering the effect of longer vocal folds with larger vibrational amplitudes on voice physiology in men (cf. 5). As compared to women, the higher mean closed quotient values in men might be due to differences in prephonatory glottal width, as well as thicker vocal folds. The minimum flow was also observed to be significantly higher in men. This finding conflicts with the observed better closure of the vocal folds in men (3). It probably can be explained by the algorithm for labelling the moments of closure and opening of the vocal folds, which might introduce the inclusion of the "piston flow" in the minimum flow (cf. 26,35,41). This "piston flow" is observed during the process of vertical phasing of especially male vocal folds and, thus, leads to higher minimum flows in men.

The findings of the present study show differences in GVVW between men and women, which are highly consistent with those reported in the literature (11,12,18,27,41). The differences in GVVW parameters also imply the presence of specific gender-related voice source spectra (46). Apart from the fundamental frequency, these characteristics of the voice source spectrum might be the basis for a perceptual distinction between gender.

Vocal training.To our knowledge the present study is the first one in which differences in GVVWs are investigated in large groups with and without vocal training. The statistical analysis showed a few differences between subjects with and without vocal training.

In the women with vocal training, larger mean closing quotients and MFDRs were found as compared to the untrained subjects. An increase in both the closing quotient and MFDR seems contradictory, considering a higher closing velocity of the vocal folds in a longer time period. An explanation might be found in the higher ac flow values of trained subjects compared to the untrained ones. From studies investigating the effect of singing style on the voice source, it is well known that singers use less adductory forces, presumably resulting in larger vocal fold amplitudes and higher ac flows (24,29,30,34, 55,63). Larger amplitudes of vocal fold vibration may well lead to higher ac flow and MFDR values and larger closing quotients.

The only significant difference between the trained and untrained male subjects was for the speed quotient, with the trained subjects having larger values and, thus, a more asymmetrical pulse shape. Skewing of the waveform is, among others, determined by the relative difference between the duration of the opening and the duration of the closure of the vocal folds. The duration of the closure is influenced by factors such as Bernoulli forces. The more pronounced skewing of the waveform to the right in trained men might be caused by the slightly higher ac flow levels (p=0.045; see Table 3), glottal geometry (64) or by source filter interaction (inertia of the vocal tract) (50).

Knowledge about glottal function in trained persons was expected to give directions towards the definition of potentially "good" vocal behaviour. However, only a few differences in the parameters were found between trained and untrained subjects.

In the present study, the phonation tasks did not include specific singing tasks, which might have revealed differences between untrained and trained subjects in the sensorineural motor control of the voice source (33,65), possibly resulting in more pronounced differences in GVVW.

Sound pressure level and intra oral pressure. Apart from the influence of intensity conditions (soft, normal, loud) on GVVW characteristics, which has already been discussed in the previous sections, the influence of SPL was also investigated within the subgroups. Twelve subgroups were created according to gender, status of vocal training and intensity condition. The influence of SPL on GVVW parameters is very clear (see Tables 3 and 4). There is a strong positive relationship between subglottal pressure (P_{sub}) and SPL. An increment in P_{sub} is explained both by an increase of glottal flow and glottal resistance. The increase in flow leads to higher ac flows and, in specific conditions, to higher average flows, while on the other hand the increase in adductory forces, expressed by the glottal resistance, induces an increase in closed quotient and a decrease in minimum flow. These combined effects produce varying closing quotients, decreasing speed quotients in men and exponentially increasing MFDR values. While an increment in SPL is positively related to the closed quotient, P_{sub} itself, as the force that drives the vocal folds apart, has a negative relation with the closed quotient in men.

CONCLUSIONS

Replication of data is important to provide verification of suggested processes or hypothesized functioning of systems (66). In this study glottal functioning was investigated in relation to the factors vocal training, gender and intensity condition, and the variables SPL, fundamental frequency, IOP, and age.

Compared to previous investigations, this study found higher average absolute values of the flow-based parameters ac flow and MFDR. These differences are related to the higher mean SPL values measured in the present study, as compared to previous studies. Other reasons for the observed differences might be the phonatory task contents, equipment or parameter extraction algorithms. Instead of using a word or a sentence, as in the present study, most of the previous studies employed the CVC sequence /bæpbæp/ to acquire glottal flow characteristics. To avoid bias influence of task content, uniformity in phonation tasks is recommended in order to facilitate the use of exchangeable data bases for frames of reference. Before universally accepting either set of phonation tasks, it should be thoroughly analyzed for its representability of normal vocal fold function. Another concern deals with equipment. Previous investigations revealed an important influence of low pass filter settings on resulting parameter values. Low pass filter settings should be standardized to facilitate making comparisons between investigations. In this respect, finally, parameter extraction should also be performed according to widely accepted guidelines.

From the results of the present study we conclude that voice function basically does not differ between subjects with vocal training and untrained subjects; however differences do exist regarding opening and closing of the vocal folds in a glottal cycle, as well as the velocity of these dynamic events. A phonation task illustrating appropriate singing abilities of the voice source in trained subjects, however, was missing. Such a task might have revealed voice source adjustments which were not obtained in the present study. Because of no differences in flow-based parameters and closed quotient between the trained and the untrained subjects, no statement can be made about a possible "good" vocal behaviour. By way of epidemiologic definition, abnormal voice production, however, is reflected in deviations from normative values of these parameters. Depending on the direction of the deviations, hyperfunctional or hypofunctional voicing can be observed. More glottal flow measurements should be performed to make a further differentiation within these main categories possible.

A number of differences were established between glottal waveform characteristics of men and women. The observed differences can be explained by gender-related differences in anatomical constitution. Differences between men and women also exist in their strategy for varying SPL and F_0 by adjusting phonatory mechanisms, as reflected in GVVW characteristics. The differences in GVVW between men and women also provide a possible distinction in voice source spectrum, which makes speech, apart from pitch, characteristic for men and women.

References

1. Sulter A, Schutte H, Miller D. Differences in phonetogram features between male and female subjects with and without vocal training. J Voice 1995;9:363-77.
2. Sulter A, Schutte H, Miller D. Standardized Laryngeal Videostroboscopic Rating: Differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency and age. J Voice 1996;10:175-89.
3. Sulter A, Albers F. The effects of frequency and intensity level on glottal closure in normal subjects. Clin Otolaryngol 1996:in press.
4. van den Berg Jw, Zantema J, Doornenbal P. Jr. On the air resistance and the Bernoulli effect of the human larynx. J Acoust Soc Am 1957;29:626-31.
5. Titze I, Talkin D. A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. J Acoust Soc Am 1979;66:60-74.
6. Story B, Titze I. Voice simulation with a body-cover model of the vocal folds. J Acoust Soc Am 1995;97:1249-60.
7. Miller R. Nature of the vocal cord wave. J Acoust Soc Am 1959;31: 667-77.
8. Fritzell B. Inverse filtering. J Voice 1992;6:111-4.

9. Miller C, Daniloff R. Airflow measurements: theory and utility of findings. J Voice 1993;7:38-46.
10. Rothenberg M. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. J Acoust Soc Am 1973;53:1632-45.
11. Holmberg E, Hillman R, Perkell J. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. J Acoust Soc Am 1988;84:511-29.
12. Perkell J, Hillman R, Holmberg E. Group differences in measures of voice production and revised values of maximum airflow declination rate. J Acoust Soc Am 1994;96:695-8.
13. Karlsson I. Glottal wave forms for normal female speakers. STL-QPSR 1985;1:31-6.
14. Lindqvist J. The voice source studied by means of inverse filtering. STL-QPSR 1970;1:3-9.
15. Sundberg J, Gauffin J. Waveform and spectrum of the glottal voice source. In: Lindblom B, Öhman S, eds. Frontiers of speech communication research. London: Academic Press, 1979:301-20.
16. Stathopoulos E, Sapienza C. Respiratory and laryngeal measures of children during vocal intensity variation. J Acoust Soc Am 1993;94:2531-43.
17. Sapienza C, Stathopoulos E. Comparison of maximum flow declination rate: Children versus adults. J Voice 1994;8:240-7.
18. Higgins M, Saxman J. A comparison of selected phonatory behaviors of healthy aged and young adults. J Speech Hear Res 1991;34:1000-10.
19. Colton R, Brewer D, Rothenberg M. Noninvasive evaluation of voice. In: Kirikae I, ed. Proceedings of the XXth Int. Congr. Logoped. Phoniatr. IALP. Organizing Committee of the XXth Congr IALP, Tokyo, 1986:322-3.
20. Hillman R, Holmberg E, Perkell J, Walsh M, Vaughan C. Objective assessment of vocal hyperfunction: an experimental framework and initial results. J Speech Hear Res 1990;32:373-92.
21. Hillman R, Holmberg E, Perkell J, Walsh M, Vaughan C. Phonatory function associated with hyperfunctionally related vocal fold lesions. J Voice 1990;4:52-63.
22. Sulter A, Schutte H. On the existence of speaker-specific maximum flow declination rate (MFDR) - sound pressure level (SPL) profiles. J Acoust Soc Am 1992;91:2420.
23. Sulter A, Schutte H. Differences in parameterised Glottal Volume Velocity Waveforms between and within speaker groups. Proceedings of the second International Symposium Phonosurgery (Amsterdam), 1992:57.
24. Peterson K, Barkmeier J, Verdolini-Marston K, Hoffman H. Comparison of aerodynamic and electroglottographic parameters in evaluating clinically relevant voicing patterns. Ann Otol Rhinol Laryngol 1994;103:335-46.
25. Sapienza C, Stathopoulos E. Respiratory and laryngeal measures of children and women with bilateral vocal fold nodules. J Speech Hear Res 1994;37:1229-43.
26. Dromey C, Stathopoulos E, Sapienza C. Glottal airflow and electroglottographic measures of vocal function at multiple intensities. J Voice 1992;6:44-54.
27. Stathopoulos E, Sapienza C. Respiratory and laryngeal function of women and men during vocal intensity variation. J Speech Hear Res 1993;36: 64-75.
28. Colton R. Glottal waveform variations associated with different vocal intensity levels. In: Lawrence van L, ed. Transcr. XIIIth Symp Care of the Professional Voice. The Voice Foundation, New York, 1984:39-45.
29. Sundberg J, Gramming P, Lovetri J. Comparisons of pharynx, source, formant, and pressure characteristics in operatic and musical theatre singing. J Voice 1993;7:301-10.
30. Hertegård S, Gauffin J, Sundberg J. Open and covered singing as studied by means of fiberoptics, inverse filtering, and spectral analysis. J Voice 1990;4: 220-30.
31. Shipp T. Izdebski K. Vocal frequency and vertical larynx position by singers and nonsingers. J Acoust Soc Am 1975;58:1104-6.
32. Watson P, Hixon T. Respiratory kinematics in classical (opera) singers. J Speech Hear Res 1985;28:104-21.
33. Teachey J, Kahane J, Beckford N. Vocal mechanics in untrained professional singers. J Voice

1991;5:51-6.
34. Sundberg J, Rothenberg M. Some phonatory characteristics of singers and nonsingers. STL-QPSR 1986;4:65-77.
35. Cranen B, Boves L. On the measurement of glottal flow. J Acoust Soc Am 1988;84:888-900.
36. Shipp T. Intra oral air pressure and lip occlusion in midvocalic stop consonant production. J Phonetics 1973;1:167-70.
37. Rothenberg M. Measurements of airflow in speech. J Speech Hear Res 1977;20:155-76.
38. Smitheran J, Hixon T. A clinical method for estimating laryngeal airway resistance during vowel production. J Speech Hear Dis 1981;46:138-46.
39. Löfqvist A, Carlborg B, Kitzing P. Initial validation of an indirect measure of subglottal pressure during vowels. J Acoust Soc Am 1982;72:633-5.
40. Lindqvist J. Studies of the voice source by means of inverse filtering. STL-QPSR 1964;4:8-13.
41. Hertegård S, Gauffin J, Karlsson I. Physiological correlates of the inverse filtered flow waveform. J Voice 1992;6:224-34.
42. Hertegård S, Gauffin J. Acoustic properties of the Rothenberg mask. STL-QPSR 1992;2-3:9-18.
43. Schutte H. The efficiency of voice production. Ph.D. dissertation, University of Groningen, 1980.
44. Hair J. Jr., Anderson R, Tatham R, Black W. Multivariate data analysis. New York: MacMillan Publishing Company, 1992:153-92.
45. Tielen M. Male and female speech. An experimental study of sex-related voice and pronunciation characteristics. Thesis, Universiteit van Amsterdam, 1992.
46. Monsen R, Engebretson A. Study of variations in the male and female glottal wave. J Acoust Soc Am 1977;62:981-93.
47. Higgins M, Saxman J. Inverse-filtered air flow and EGG measures for sustained vowels and syllables. J Voice 1993;7:47-53.
48. Koreman J. Personal communication. 1995.
49. Till J, Jafari M, Crumley R, Law-Till C. Effects of initial consonant, pneumotachographic mask, and oral pressure tube on vocal perturbation, harmonics-to-noise, and intensity measurements. J Voice 1992;6:217-23.
50. Rothenberg M. Acoustic interaction between the glottal source and the vocal tract. In: Stevens K, Hirano M, eds. Vocal fold physiology. Tokyo: U Tokyo Press, 1981:305-28.
51. Alku P, Vilkman E. Effects of bandwidth on glottal airflow waveforms estimated by inverse filtering. J Acoust Soc Am 1995;98:763-7.
52. Brodnitz F. Contact ulcer of the larynx. Arch Otol 1961;74:70-80.
53. Kent R, Kent J, Rosenbek J. Maximum performance tests of speech production. J Speech Hear Dis 1987;52:367-87.
54. Tanaka S, Gould W. Relationships between vocal intensity and noninvasively obtained aerodynamic parameters in normal subjects. J Acoust Soc Am 1983;73:1316-21.
55. Gauffin J, Sundberg J. Data on the glottal voice source behavior in vowel production. STL-QPSR 1980;2-3:61-70.
56. Gauffin J, Sundberg J. Spectral correlates of glottal voice source waveform characteristics. J Speech Hear Res 1989;32:556-65.
57. Holmberg E, Hillman R, Perkell J, Gress C. Relationships between intra-speaker variation in aerodynamic measures of voice production and variation in SPL across repeated recordings. J Speech Hear Res 1994;37:484-95.
58. Åkerlund L, Gramming P. Average loudness level, mean fundamental frequency, and subglottal pressure: Comparison between female singers and nonsingers. J Voice 1994;8:263-70.
59. Netsell R, Lotz W, DuChane A, Barlow S. Vocal tract aerodynamics during syllable productions: Normative data and theoretical implications. J Voice 1991;5:1-9.
60. Hoit J, Hixon T. Age and laryngeal airway resistance during vowel production in women. J Speech Hear Res 1992;35:309-13.

61. Kahane J. A morphological study of the human prepubertal and pubertal larynx. Am J Anat 1978;151:11-20.
62. Titze I. Physiologic and acoustic differences between male and female voices. J Acoust Soc Am 1989;85:1699-1707.
63. Sundberg J, Titze I, Scherer R. Phonatory control in male singing: a study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. J Voice 1993;7:15-29.
64. Titze I. Parameterization of the glottal area, glottal flow, and vocal fold contact area. J Acoust Soc Am 1984;75:570-80.
65. Murry T, Caligiuri M. Phonatory and nonphonatory motor control in singers. J Voice 1989;3:257-63.
66. Muma J. The need for replication. J Speech Hear Res 1993;36:927-30.
67. Titze I. Principles of voice production. Englewood Cliffs: Prentice-Hall, 1994.

5

The clinical relevance of the relation between Maximum Flow Declination Rate and Sound Pressure Level in predicting vocal fatigue

Arend M. Sulter

Department of Otorhinolaryngology, University Hospital Groningen, Groningen, the Netherlands

Submitted

INTRODUCTION

Intensive voice use, that is speaking for a long time, or speaking at high sound intensities, may lead to vocal problems (1). The etiology of the qualitative deterioration of the voice has been described as vocal fatigue (2). Although susceptibility to vocal fatigue has been investigated, a clear picture of circumstances leading to vocal problems is still lacking (1-3). In some persons high demands on the vocal apparatus appear to provoke myogenic insufficiency of the muscles involved in voice production (4) while in others mechanical forces acting on vocal fold tissue may be responsible for the development of edema (5-6). Histological investigations of vocal fold tissue with benign pathology show abnormalities suggestive of repetitive traumata to the vocal fold cover (7-10).

Trauma of the vocal fold cover can be caused by colliding vocal folds which occurs cyclically during phonation. An indicator of vocal fold collision forces is Maximum Flow Declination Rate (MFDR) (11-12). MFDR can be measured as the minimum of the first derivative of the Glottal Volume Velocity Waveform (Figure 1). MFDR has an exponential relationship with Sound

Pressure Level (SPL) (13), a relationship which implies nonlinear increment of colliding forces with increasing SPL (Figure 2). An important feature of MFDR is its systematic relationship with SPL (13-15). A specific intra-individual relationship between SPL and MFDR, and thus collision forces, should have clinical consequences regarding the susceptibility to vocal fatigue and vocal fold pathology (16). The specific individual relationship between MFDR and SPL will hereafter be reffered to as the Analytic MFDR-SPL (AMS) curve.

The AMS curve can be described mathematically with the equation (13,16):

(1) $MFDR = a + b^{SPL}$

The intercept a is the constant in the equation and represents the maximum decrease of glottal flow rate at an intensity level of 60 dB. At higher intensity levels the glottal flow rate decreases more abruptly, causing greater pressure variations. This increase in MFDR is determined by base b.

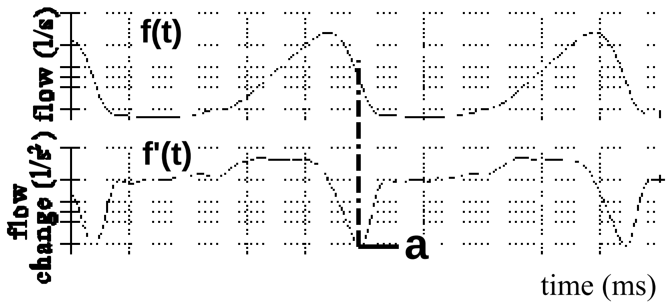


Figure 1.Glottal Volume Velocity Waveform (above) with its first derivative (below). The minimum of the first derivative (a) gives the maximum flow declination rate and represents the closing velocity of the vocal folds.

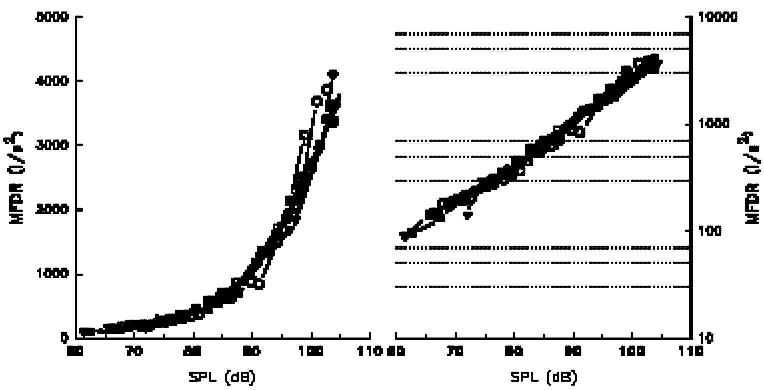


Figure 2.Relation between Maximum Flow Declination Rate (MFDR) and Sound Pressure Level (SPL) for two male subjects (hollow and filled symbols, respectively), phonating at multiple intensities. Left metric scales, right logarithmic y-axis. Figures show the exponential relation between SPL and MFDR.

Although many attempts have been made to diagnose susceptibility to vocal fatigue, a test revealing a causative relationship between vocal demand and vocal fatigue has not yet been accepted. The nature of the specific intra-individual relationship between SPL and MFDR might function as such a clinical tool. To test the hypothesis that differences exist between subjects regarding their AMS curves, measurements were performed in groups of subjects with differing susceptibility to vocal fatigue.

METHODS

Definition of vocal fatigue

Subjects invited to participate in this study were asked to rate their own susceptibility to vocal fatigue. Vocal fatigue was defined subjectively as an inability to respond to vocal demands in combination with a decrease of vocal dynamic ranges (pitch and intensity) (1). Inability to respond to vocal demands was rated on a five point scale ranging from full adequacy to respond to vocal demands under all circumstances, to an insufficient load tolerance of the voice in normal daily use. Vocal dynamics were rated on a four point scale, ranging from good to poor.

Subjects

Out of 100 adults three main groups were created, according to self-rated susceptibility to vocal fatigue (Table 1). Thirteen patients (9 females, 4 males) with videolaryngostroboscopically confirmed vocal fold pathology served as

	male (mean age; SD)	female (mean age; SD)
Vocal fold pathology (n=13)	4 (27.5; 12.82)	9 (32.4; 10.99)
Subjects without complaints (n=44)	18 (24.6; 5.10)	26 (23.6; 11.05)
Trained group (n=43)	25 (38.4; 17.16)	18 (30.8; 13.47)
Total (n=100)	47 (32.2; 14.82)	53 (27.6; 12.34)

Table 1. Number and age (mean and standard deviation [SD]) of subjects in groups.

one group. They all scored maximally negative on the scales and, therefore, showed the highest possible susceptibility to vocal fatigue. The vocal fold pathologies in the nine female patients (mean age 32.4, standard deviation [SD] 10.99 years) consisted of eight cases of vocal fold nodules or broad-based swellings and one case of a submucosal cyst. Pathology in the four male patients (mean age 27.5, SD 12.82 years) consisted of two cases of a sulcus vocalis, one case of vocal fold nodules and one case of a submucosal cyst.
Twenty-six female (mean age 23.6 years, SD 11.05) and 18 male (mean age 24.6 years, SD 5.10) subjects without vocal complaints were used as a control group. Freedom from vocal fold pathology was established laryngostroboscopically.

Vocal characteristics of subjects with voice training may differ from those of subjects without vocal training (3). Therefore, a third group was created consisting of amateur singers with a minimum of 2 years of vocal training (17). The mean age of 18 female subjects was 30.8 years (SD 13.47), and the mean age of 25 male subjects was 38.4 years (SD 17.16).

Speech material

The Dutch word stagiaire, /stazjærœ/ (trainee)and the Dutch sentence hou eens op te blèren, /hou ens op te blæren/ (stop bawling) were produced at three sound intensity levels, namely, soft, comfortable (hereafter referred to as normal), and loud by each subject. The intensities were chosen by the subject with the investigator's approval and excluded whispering and shouting. Subjects were instructed to prolong the /æ/ vowel during each production. They were permitted to chose the intensity level to assure the most natural voice production. This intended condition was further facilitated by using an /æ/ vowel in a word and a sentence.

Equipment

Glottal Volume Velocity Waveforms (GVVW) were acquired using a circumferentially vented pneumotachograph, also known as the Rothenberg mask (18), in combination with the Glottal Enterprises MS-100 system with inverse filtering MSIF-2 units and a 14-bit digital memory (Cutec CD-425). A 400 ms portion of the oral flow signal originating from the MS-100 unit was stored in the digital memory by activating a hold circuitry with a foot switch. The stored oral flow signal was read out repetitively to the inverse filtering unit. GVVWs were obtained by removing resonance effects by inverse filtering.

The audio signal was recorded with a Sennheiser Back-Elektret-Kondensator-Microphone MKE 2 mounted on the Rothenberg mask at a fixed distance of 7 cm from the mouth. Sound pressure level (SPL) was derived from the audio signal with an integration time of 100 ms. A dB(A) filter was used to exclude background noise. The placement of the mask between mouth and microphone resulted in a 5 dB attenuation of the acoustic signal.

Inverse filtering

To compensate for the resonances of the vocal tract, the 400 ms digitally stored oral flow signal was manually inverse filtered with the MSIF-2 unit. The process of filtering was performed interactively by visualizing the result of adjusting two frequency selective attenuators on an oscilloscope (Hameg Digital Storage Scope HM 208). The goal of adjusting the filters was to arrive at a maximally flat portion of that part of the GVVW which represents the closed phase of the glottal cycle (19). To remove high frequency noise, the derived glottal flow signal was low pass filtered (Frequency Devices 8 pole

Bessel 902 LPF) with a cut-off frequency of 1.6 kHz, consistent with the resonance characteristics of the mask (19).

Signal registration

The glottal flow signal and the Sound Pressure Level signal were registered on VHS tapes with an instrumentation recorder (TEAC XR-510 cassette data recorder) at a speed of 38.1 cm/s, offering an effective frequency range from DC to 10 kHz.

Calibration

Before each measurement session the equipment was calibrated for flow and sound pressure level. The mask with the pressure transducer was calibrated at 0, 400 and 800 ml/s air flow rates, by placing the mask with a tight seal against an artificial head that had a laminar flow connection with a central air supply. The exact flow level could be adjusted by means of a Brooks 2-tube sho-rate flow meter.

The sound pressure level was calibrated at 70, 75 and 80 dB by placing the mask on a mould which incorporated the B&K Artificial Voice Type 4219. The artificial voice was driven by the B&K Beat Frequency Oscillator Type 1022 at a frequency of 150 Hz.

Data acquisition

Subjects were asked to push the mask firmly against the face and to explore during expiration the possibility of undesirable leakage of air where the mask contacted the face. In a number of females, the nose was too small to properly seal against the mask. In those cases that part of the mask was filled with a mouldable silicone based impression material (Optosil P plus; Bayer Dental). With this adaptation no further leakage problems were encountered.

Recording began prior to the onset of each utterance and ended after activating the hold-circuitry. The hold switch was activated after the beginning of production of the /æ/ vowel and a 400 ms midvocalic oral flow signal was stored in the digital memory for inverse filtering and subsequent determination of the glottal flow signal. The stored signal was checked for a steady state appearance by visually comparing the level of the waveform peaks on the oscilloscope. The filters were adjusted manually to remove the formant ripple. A recording was made of the optimally corrected GVVW after completion of the filtering procedure. All examinations were performed by the same investigator.

Signal processing and data analysis

The recorded signals were digitized with a 12 bit successive approximation converter (MetraByte DASH 16). A sampling frequency of 500 Hz was used to convert calibration and speech signals. The inverse filtered signals were digitized with a sampling frequency of 10 kHz. All signals were digitized simultaneously and then demultiplexed.

A parameter extraction program was written in a fourth generation signal analysis language (ASYST, MacMillan Software Company) to analyze SPL and GVVW parameters. Calibration files created specifically for each subject were used to quantify matching signals.

A peak-picking algorithm was used to identify the fundamental period T for measurement of the fundamental frequency. MFDR was measured within each fundamental period as the minimum of the first derivative of the glottal waveform (Fig. 1).

Curve fitting

A curve fitting procedure of the SigmaPlot™ software (Jandel Corporation) was used to fit individual MFDR-SPL data points to equation 1. Resulting values for intercept a and base b were used for further statistical evaluation.

Statistical analysis

Analysis of variance (ANOVA) with post-hoc Least-Significant Differences (LSD) (SPSS Inc.) were used to investigate differences among groups (21). Intercept a and base b of equation 1 were regarded as dependent variables. Gender and vocal fatigue group were introduced as factors. In case of significant interaction between factors, oneway analysis of variance was performed at each factor level. A probability level $\alpha=0.05$ was used to test the null hypothesis that there were no differences among the variables under investigation.

RESULTS

The results of fitting individual MFDR-SPL data points to equation 1 are summarized in table 2. Men had statistically significant higher intercept values than women [F(1,98)=89.88, p<0.001]. In male subjects, vocal folds were, therefore, assumed to close at a higher velocity at 60 dB than in females. Base values were also significantly higher in men [F(1,98)=5.07, p=0.027], implying a more rapid increase of MFDR with increasing SPL, as compared to women. Because of a statistically significant interaction between factors gender and vocal fatigue group for base b [F(2,97)=3.28, p=0.042], differences between vocal fatigue groups were separately analyzed for each gender with oneway analysis of variance and post-hoc LSD tests. In men no differences were found among the groups. In women, however, a significant difference (p<0.05) was found for both intercept a and base b between the group most susceptible to vocal fatigue and both the other groups (normal and trained). The group susceptible to vocal fatigue had higher average intercept and base values which means that higher closing velocities were observed in women susceptible to vocal fatigue.

	intercept a		base b	
	men	women	men	women
vocal fold pathology	561 (159.2)	388 (125.0)	1.183 (0.0320)	1.199 (0.0124)
normal	580 (153.4)	312 (93.7)	1.187 (0.0188)	1.180 (0.0296)
vocal training	584 (180.8)	301 (99.9)	1.199 (0.0209)	1.179 (0.0163)

Table 2. Mean and standard deviation (between brackets) of intercept a and base b of the equation MFDR = a + b ^{SPL} are given according to gender for a group susceptible to vocal fatigue (vocal fold pathology) and two control groups (normal subjects without vocal complaints and a group having received vocal training).

DISCUSSION AND CONCLUSIONS

In this study differences in average intercept and base values of the equation describing AMS curves were found between on the one hand female subjects susceptible to vocal fatigue and those with a higher vocal load tolerance (normal and trained subjects) on the other. Figure 3 shows the AMS curves for the group of female subjects susceptible to vocal fatigue and the group of normal female subjects, as well as the difference in MFDR between

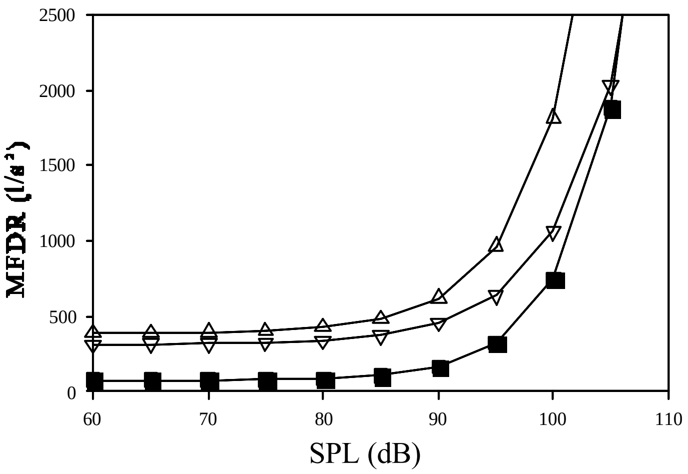


Figure 3. Reconstructed mean Maximum Flow Declination Rate (MFDR) - sound pressure level (SPL) relationships for females susceptible to vocal fatigue (△) and normal females (▽). A separate curve gives the difference (■).

the groups. dB after which it rapidly increases. This means that especially above 85 dB the flow decreases more rapidly in women susceptible to vocal fatigue than normal women. The decrease of air flow is related to the closing velocity of the vocal folds and hence to the events that occur when the vocal folds close (11,12). The collision forces and the concomitant abrupt pressure variations acting on the vocal fold tissue are in part determined by the closing velocity of the vocal folds. Several studies have shown the presumed effects of repetitive trauma on vocal fold tissue (7-10). Disturbances of the architecture of the basement membrane zone are combined with fluid accumulation (edema) and deposition of organic material in this area. Edema leads to deterioration of voice quality (22) and affects the soft intensity range (23), while deposition of organic material produces vocal fold nodules, leading to incomplete glottal closure and hence to breathiness (24).

No differences in intercept and base values were found among the male groups. This might be due to the fact that the male group susceptible to vocal fatigue consisted only of four subjects. Furthermore, two of these had a sulcus vocalis, a different category of pathology from those observed in the other patients (biomechanical forces applied to vocal folds with a sulcus are likely to result in tissue reactions different from those observed in vocal folds with swellings).

Although MFDR acts as an indicator of vocal fold collision forces, its influence on the condition of vocal fold tissue will be determined in correlation with other factors. Fundamental frequency gives the number of collisions per second. Damage of vocal fold tissue will be determined by a cumulative amount of forces applied during a certain time period. As women phonate at fundamental frequencies almost twice as high as men, the observed differences in intercept values between men and women are in this respect compensated by pitch. Following these considerations, phonation at higher pitches is potentially problematical for vocal hygiene.

Another factor determining the effect of biomechanical forces on vocal fold tissue is glottal geometry. Studies have suggested that incomplete closure of the vocal folds at the dorsal glottis provides a situation more sensitive to the development of pathology of the vocal folds (25). Localized tearing forces (26) and pressure changes (27) could be contributing factors for this pathology. Incomplete dorsal closure is encountered more frequently in females (17), which might explain the higher occurrence of vocal problems in females.

This study shows the potential for AMS curves to predict susceptibility to vocal fatigue. Given the more pronounced differences between groups at higher sound pressure levels, it might also explain why routine tests comprising prolonged reading or phonating at low intensity levels are not indicative of susceptibility to vocal fatigue. A possible source of bias in this study might be found in the presence of pathology in the group susceptible to vocal fatigue. A longitudinal study analyzing the condition of vocal fold tissue in groups with different AMS curves should confirm the clinical importance of the observations made in the present study.

References

1. Kitch J, Oates J. The perceptual features of vocal fatigue as self-reported by a group of actors and singers. J Voice 1994;8:207-14.
2. Sander E, Ripich D. Vocal fatigue. Ann Otol Rhinol Laryngol 1983;92:141-5.
3. Pausewang Gelfer M, Andrews M, Schmidt C. Effects of prolonged loud reading on selected measures of vocal function in trained and untrained singers. J Voice 1991;5:158-67.
4. Jackson C. Myasthenia laryngis. Arch Otolaryngol 1940;32:434-63.
5. Scherer R, Titze I, Raphael B, Wood R, Ramig L, Blager R. Vocal fatigue in a trained and an untrained voice user. In: Baer T, Sasaki C, Harris K, eds. Laryngeal function in phonation and respiration. Waltham, Massachusetts: College-Hill Press, 1986:533-55.
6. Titze I. Vocal fatigue: some biomechanical considerations. In: Lawrence V, ed. Transcripts of the Twelfth Symposium: care of the professional voice. NewYork: Voice Foundation, 1983: 97-104.
7. Gray S. Basement membrane zone injury in vocal nodules. In: Gauffin J, Hammarberg B, eds. Vocal fold physiology. Acoustical, perceptual, and physiological aspects of voice mechanisms. San Diego: Singular Publishing Group, Inc., 1991:21-7.
8. Dikkers F, Hulstaert C, Oosterbaan J, Cervera-Paz F. Ultrastructural changes of the basement membrane zone in benign lesions of the vocal folds. Acta Otolaryngol 1993;113:98-101.
9. Gray S, Hammond E, Hanson D. Benign pathologic responses of the larynx. Ann Otol Rhinol Laryngol 1995;104:13-8.
10. Dikkers F, Nikkels P. Benign lesions of the vocal folds: histopathology and phonotrauma. Ann Otol Rhinol Laryngol 1995;104:698-703
11. Hillman R, Holmberg E, Perkell J, Walsh M, Vaughan C. Objective assessment of vocal hyperfunction: an experimental framework and initial results. J Speech Hear Res 1989;32:373-92.
12. Hillman R, Holmberg E, Perkell J, Walsh M, Vaughan C. Phonatory function associated with hyperfunctionally related vocal fold lesions. J Voice 1990;4:52-63.
13. Sulter A, Schutte H. Glottal flow parameters in speech modes [Abstract]. ASHA 1991;33:168.
14. Gauffin J, Sundberg J. Data on the glottal voice source behavior in vowel production. STL-QPSR 1980;2-3:61-70.

15. Holmberg E, Hillman R, Perkell J, Gress C. Relationships between intra-speaker variation in aerodynamic measures of voice production and variation in SPL across repeated recordings. J Speech Hear Res 1994;37:484-95.
16. Sulter A, Schutte H. On the existence of speaker-specific maximum flow declination rate (MFDR) - sound pressure level (SPL) profiles [Abstract]. J Acoust Soc Am 1992;91:2420.
17. Sulter A, Schutte H, Miller D. Standardized laryngeal videostroboscopic rating: Differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency and age. J Voice 1996;10:175-89.
18. Rothenberg M. A new inverse-filtering technique for deriving the glottal air flow waveform during voicing. J Acoust Soc Am 1973;53:1632-45.
19. Lindqvist J. Studies of the voice source by means of inverse filtering. STL-QPSR 1964;4:8-13.
20. Hertegård S, Gauffin J. Acoustic properties of the Rothenberg mask. STL-QPSR 1992;2-3:9-18.
21. Hair J Jr., Anderson R, Tatham R, Black W. Multivariate data analysis. New York: MacMillan Publishing Company, 1992.
22. Titze I, Talkin D. A theoretical study of the effects of various laryngeal configurations on the acoustics of phonation. J Acoust Soc Am 1979;66:60-74.
23. Titze I, Jiang J, Hsiao T. Measurement of mucosal wave propagation and vertical phase difference in vocal fold vibration. Ann Otol Rhinol Laryngol 1993;102:58-63.
24. Lancer J, Syder D, Jones A, Le Boutillier A. Vocal cord nodules: a review. Clin Otolaryngol 1988;13,43-51.
25. Morrison M, Rammage L, Belisle G, Pullan C, Nichol H. Muscular tension dysphonia. J Otolaryngol 1983;12:302-6.
26. McGowan R. An analogy between the mucosal waves of the vocal folds and wind waves on water. Haskins Laboratories Status Report on Speech Research 1990;101/102:243-9.
27. Story B, Titze I. Voice simulation with a body-cover model of the vocal folds. J Acoust Soc Am 1995;97:1249-60.

6

Perceptive characteristics of speech of untrained and trained subjects, and influences of gender and age

Arend M. Sulter and Herman F. Peters*

Department of Otorhinolaryngology, University Hospital Groningen, Groningen, and * Department of Otorhinolaryngology, University Hospital Nijmegen, Nijmegen, the Netherlands

Submitted

INTRODUCTION

This study is the result of an extensive research project, originated to supply information on normal voice production and to create a frame of reference for qualifying vocal performance. Apart from a large group of subjects without vocal complaints or vocal pathology, another group of subjects having received vocal training and regularly exploiting their vocal abilities, was investigated by different means, to give direction to what should be regarded as good vocal performance in the continuum from poor to excellent. Related studies have already established normative data regarding phonetograms (1), laryngostroboscopy (2), and glottal closure (3). In this article speech characteristics of subjects are evaluated.

The complex acoustic event representing speech results from a basic voice source signal and the modulation of this signal by the articulatory organs. One important method to qualitatively analyze speech is perceptual evaluation (4). Although listeners vary in their qualitative description of speech sounds (5-7), a structured approach using a multidimensional scaling instrument provides reliable perceptual judgments on running speech with constantly emerging factorial dimensions (8,9).

Previous studies resulting from our research project concentrated on characteristics of the voice source. Differences between men and women were observed in melodic and intensity ranges (1), in both laryngeal appearance and glottal functioning (2,3), as well as in vocal fold physiology (4). Less appealing differences were noted between trained and untrained subjects; however, results of phonetography indicated that trained subjects might benefit from a superior control over the voice source (1).

Speech has been the subject of many investigations, concentrating on different aspects, amongst others perception of speech. However, limited information is available on speech characteristics, as determined with a standardized scaling instrument in large groups of subjects. Using a standardized scaling instrument has several advantages. It employs the tedious work of determining scales that can be utilized in a rating experiment. An instrument with selected scales forces judges to express their opinion in a standardized way, thus excluding incompatible variety of expressions. Using a specific scaling instrument also offers the possibility of comparing results between investigations. One of the few studies using a standardized scaling instrument on evaluating laryngeal speech of large groups of men and women was conducted by Tielen (9). Only a few distinct differences were determined between men and women. Compared to men, women were evaluated to speak with higher pitch and more melodious. The specific choice of subjects, in relation with their profession, might have caused the resemblance in speech characteristics between men and women. Other cohorts of men and women might have revealed a different picture, more close to what is expected from literature. The scaling instrument employed by Tielen (9) was modified to concentrate on differences in sociocultural aspects of speech. Scales more specifically reflecting physiology of phonation had a less prominent place, although this information could be used to relate perceptual aspects of speech with a physiologic basis to both quality of speech and physiologic measurements.

In previous studies trained subjects showed a possible superior control over the voice source (1,11). Perceptual evaluation of speech of trained subjects might demonstrate effects of this superior control.

With the former considerations, the present study investigates aspects of speech and related vocal and articulatory processes, to answer the following questions:

1. Can speech of a large number of voice healthy subjects reliably be evaluated with the perceptual scaling instrument of Fagel et al. (8)?

2. Does factor analysis of scale scores result in a practicable solution, fit for a further evaluation of perceptive data?

3. Do groups differ in scale scores; and if there are differences, how can they be specified according to the grouping variables gender and vocal

training?

4. Are differences between groups also reflected in differences in factor scores and what do they represent in perceptual dimensions?

5. Which scales do best differentiate between men and women, and between subjects with and without vocal training?

METHODS

Nieboer, De Graaf & Schutte (12) evaluated speech of alaryngeal speakers in a similar way we intended to do for laryngeal speech in the present study, also using --a modified version of-- the scaling instrument proposed by Fagel et al. (8). Therefore the procedural approach of Nieboer et al. (12) was used as a guideline.

Subjects

Speech samples were provided by a total of 224 Dutch untrained and trained subjects of both genders, categorized accordingly into 4 groups. The untrained subjects were recruited from groups of students and volunteers without vocal complaints or history of vocal pathology. The group consisted of 92 females and 47 males. The mean age for the female subjects in this subgroup was 20.3 years, ranging from 17 to 44 (median 19 years; standard deviation [SD] 7.37), while the mean age for the male subjects was 25.0 years, ranging from 17 to 35 (median 25 years; SD 4.68 years). Eighteen of the female, and 16 of the male subjects were smokers.

42 female and 43 male amateur singers with a minimum of two years of vocal training served as another group. The vocal training could either consist of singing in a choir that organized rehearsals with a minimum frequency of once a week, or receiving individual singing lessons with a similar minimum frequency. All choirs had a professional conductor and used auditions to admit new members. Although a minimum of 2 years of organized singing was used as a selection criterion to be included in the trained group, about 60% of the trained subjects had a considerably longer history of singing in a choir (> 5 years). The mean age of the female trained group was 35.1 years, ranging from 18 to 59 (median 34 years; SD 11.86 years), and the mean age of the male subjects was 47.5 years, ranging from 21 to 75 (median 49 years; SD 18.52 years). Five of the female, and 11 of the male trained subjects were smokers. Because all participants in this study volunteered, we refrained from matching according to age.

Scales

Speech characteristics were analyzed by way of 14 bipolar semantic scales with seven points (13), ranging from -3 to 3. The scales were taken from Fagel et al. (8), who carefully developed a scaling instrument by reducing numerous adjectives to a practicable number of so-called Alpha scales, making possible a global perceptual description of a speaker in a multidimensional perceptual space. A few changes were implemented. The scale "creaky-not creaky" was introduced to focus on a voice characteristic not incorporated in the original instrument, and the scales "dragging-brisk" and "slow-quick" were replaced by the alternative "slurring-sprightly". Instead of "husky-not husky" the scale ends "breathy-not breathy" will be used, because they reflect more closely the original Dutch terms.

The scales were all given the same direction or polarization: the more "negative" or "unfavourable" pole was placed on the left-hand side, the more "positive" or "favourable" pole on the right-hand side.

Adjectives of the scales were translated from Dutch. Minor shifts in meaning may therefore have occurred. The Dutch and English terms are listed in Table 1. Throughout this paper, English terms will be used. While referring to the results of this study, the reader is advised to give the original Dutch term with the English translation.

Listeners were also asked to estimate the age of the speaker. This information might be used to study correlations with other scales.

Scale	factor 1 eigen-value	factor 1 % var	factor 2 eigen-value	factor 2 % var	factor 3 eigen-value	factor 3 % var
1 expressionless-expressive (*expressieloos-expressief*)	12.3	45.7	1.5	5.5	1.3	5.0
2 monotonous-melodious (*monotoon-melodieus*)	12.0	44.6	1.5	5.7	1.4	5.0
3 slurring-sprightly (*slepend-levendig*)	11.2	41.5	1.5	5.7	1.3	4.9
4 shrill-warm (*schel-warm*)	12.5	46.2	1.4	5.2	0.9	3.5
5 high-low for a (wo)man (*hoog-laag voor een man/vrouw*)	12.3	45.6	1.2	4.5	1.1	4.2
6 ugly-beautiful (*lelijk-mooi*)	12.1	44.7	1.7	6.1	1.1	4.0
7 unpleasant-pleasant (*onplezierig-plezierig*)	11.9	43.9	1.7	6.4	1.3	4.6
8 breathy-not breathy (*hees-niet hees*)	11.5	42.6	1.4	5.3	1.2	4.3
9 creaky-not creaky (*krakerig-niet krakerig*)	8.1	30.1	1.7	6.2	1.4	5.3
10 dull-clear (*dof-helder*)	8.5	31.4	1.5	5.7	1.5	5.5
11 soft-loud (*zacht-luid*)	10.5	38.8	1.4	5.0	1.3	4.8
12 weak-powerful (*zwak-krachtig*)	10.2	37.9	1.3	4.9	1.3	4.6
13 broad-cultured (*plat-beschaafd*)	10.8	39.9	1.4	5.1	1.3	4.8
14 slovenly-polished (*slordig-netjes*)	9.3	34.5	1.6	6.0	1.4	5.3
15 estimated age (*geschatte leeftijd*)	22.1	81.9	0.5	1.9	0.5	1.7

Table 1. Factor analysis of the scores on 15 scales given by 27 listener judges. % var = percentage of variation. Scales are given in the original Dutch version (italics) and an English translation.

Speech samples

Speech stimuli were obtained by making high quality recordings of subjects reading a text of neutral content (De noorderwind en de zon). Equipment consisted of a B&K 4003 microphone, a B&K amplifier (type 2812) and a Sony PCM SL-F1E recorder. Recording level was adjusted for each subject to optimize signal to noise ratio. A text was preferred above spontaneous speech to control for individually differing lexicon and syntax. From the recordings three stimulus tapes were made, containing the text of about 50 seconds of speech, hereafter referred to as a sample, of each subject. Samples were randomly copied to these three tapes. The first tape started with 5 samples to have material for the judges to practise rating the scales. These scores were not used for further evaluation. The next 10 samples were also used as the last samples on the third tape in order to provide information on intrajudge reliability.

Listeners

Twenty-seven female students of the Academy for Logopedics in Nijmegen (mean age 25.0 years, standard deviation 6.07 years) in the third year of the training course rated all samples with the scaling instrument discussed. The students were regarded as naive judges because of both the fact that during their training little time was spend on perceptual evaluation of voices, as well as the limited time available to score each scale (about three seconds). Naive judges were favoured, as they are known to give adequate judgments (12) and more uniform ratings than expert raters (5), as well as time problems with expert raters, given the large numbers of samples to be judged. The presence of only female judges should not result in biased ratings (8,9). Each of the judges was given an honorarium of hfl. 40,-.

Rating procedure

The judges were instructed to rate according their first impression of the speech. To make them familiar with using the scaling instrument they were presented a randomly chosen number of five speakers.

All samples were rated in three afternoon sessions of 2 hours, each on the same day of consecutive weeks. Each session consisted of three blocks of 30 minutes rating time with in-between breaks of 10 minutes. None of the judges reported loss of concentration during the experiment.

Filtering and amplification of the sound were adjusted in such a way as to resemble as much as possible the unfiltered sound as it could be heard using headphones. The sound quality in different places in the room was checked both before and during the experiment. Care was taken to give all speech samples approximately the same loudness level when they were played to the listeners.

Data reduction

Descriptive and inferential statistical evaluation was performed with the SPSS software package (SPSS Inc.). Means and standard deviations per speaker per scale, as well as bar charts of the scores were computed. Two-way Analysis of (Co)Variance (ANCOVA) was used to determine significant effects of the grouping variables gender and vocal training, as well as the covariant age on scales and factors. If a significant interaction was present between grouping variables, separate oneway ANOVAs were performed for each grouping variable. The significance level was set at p=0.05.

Factor analysis was performed on the scores in order to gain insight into the dimensionality of the perceptual space used for judging the speakers. The factor analysis performed was a principal-component analysis with iteration. A factor solution with 6 components was pursued, five of the components being reserved for the solution found by Fagel et al. (8), and one for the estimated age. Factors with an eigenvalue < 1.0 were therefore accepted. The factors produced in this way were rotated according to the varimax procedure in order to get a clearer factor configuration (14,15). The varimax rotation does not affect the orthogonal factor structure. This means that factors are independent of each other.

For each speaker, factor scores on each factor were computed by multiplying the speaker's standardized mean score on those scales loading highest on that factor, with the corresponding factor-score coefficient. The sum of these products is the factor score. As six main factors were selected, each speaker could be characterized with the six figures representing his/her factor scores. Factors scores are calculated on the basis of the standardized scores; therefore the factor scores can be either positive or negative.

A discriminant analysis was performed on the scale means of the speakers, in order to determine which set of scales could best discriminate between men and women, and, separately for each gender, between untrained and trained subjects. The analysis was carried out according to the Rao's V method, which, in the creation of a discriminant function, selects or deletes variables on the basis of their contribution to the increase in Rao's V. Rao's V is a generalized measure of the distance between the groups along the one possible dimension.

RESULTS

Scales properties

The uniformity in using a same definition of scale ends by the listener group was checked by performing a factor analysis on all scores given by the listeners, separately for each scale. Table 1 gives the result of the factor analysis. The first three factors are given to show the structure of the factor solution. All scales have by far the highest loading on the first factor. Apart from estimated age, eigenvalues range from 12.5 to 8.1 for scales "shrill-warm" and "creaky-not creaky", respectively. The second factor shows much smaller eigenvalues ranging from 1.7 to 1.2. Therefore ratings given on scales can be regarded as given by one group with a same representation of scale ends and thus of scale use.

Table 2 shows mean correlation between raters, effective reliabilities, and minimum number of raters needed to obtain an effective reliability of 0.90 and 0.95, for the 15 scales.

$$R_e = \frac{nr_m}{(1+(n-1)r_m)} \qquad (1)$$

The "effective reliability" or "standardized item alpha" (where "item" is to be read in our case as "rater", SPSS Inc., (16)) of the 15 scales used in the rating experiment was computed according to the formula (1) where R_e is the effective reliability, n is the number of raters, and r is the mean correlation between raters. In general, high effective reliability figures were found. Values are above 0.9, with the exception of the scales "creaky-not creaky" and "dull-clear", which show reliability values of 0.884 and 0.895, respectively.

In general, ratings on a scale with an effective reliability of > 0.90 are

	Scale	Mean correlation between raters	Effective reliability	Minimum number of raters for effective reliability of 0.90	0.95
1	expressionless-expressive	0.40	0.947	14	29
2	monotonous-melodious	0.39	0.946	15	30
3	slurring-sprightly	0.35	0.935	17	36
4	shrill-warm	0.39	0.944	15	30
5	high-low for a (wo)man	0.35	0.936	17	36
6	ugly-beautiful	0.38	0.943	15	31
7	unpleasant-pleasant	0.36	0.939	16	34
8	breathy-not breathy	0.35	0.937	17	36
9	creaky-not creaky	0.22	0.884	32	68
10	dull-clear	0.24	0.895	29	61
11	soft-loud	0.33	0.931	19	39
12	weak-powerful	0.32	0.928	20	41
13	broad-cultured	0.33	0.930	19	39
14	slovenly-polished	0.26	0.907	26	55
15	estimated age	0.78	0.990	3	6

Table 2. Mean correlation between raters, effective reliability, and minimum number of raters needed to obtain an effective reliability of 0.90 and 0.95, for the 15 scales used in the rating experiment.

considered to give reliable information (17). The minimum number of raters needed to obtain a pre-defined effective reliability is calculated according to the formula

$$n_{min} = \frac{1-r_m}{r_m}\frac{R_e}{1-R_e} \qquad (2)$$

where n_{min} is the minimum number of judges, R_e is the effective reliability to be obtained, and r_m is the mean correlation between raters. For the bipolar semantic scales the minimum number of raters needed to obtain an effective reliability of 0.90 ranged from 14 ("expressionless-expressive") to 32 ("creaky-

	Scale	c.c.	p-level
1	expressionless-expressi-ve	0.91	p < 0.001
2	monotonous-melodious	0.89	p < 0.001
3	slurring-sprightly	0.93	p < 0.001
4	shrill-warm	0.89	p = 0.001
5	high-low for a (wo)man	0.91	p < 0.001
6	ugly-beautiful	0.92	p < 0.001
7	unpleasant-pleasant	0.92	p < 0.001
8	breathy-not breathy	0.91	p < 0.001
9	creaky-not creaky	0.95	p < 0.001
10	dull-clear	0.87	p = 0.001
11	soft-loud	0.91	p < 0.001
12	weak-powerful	0.92	p < 0.001
13	broad-cultured	0.89	p = 0.001
14	slovenly-polished	0.69	p = 0.028
15	estimated age	0.99	p < 0.001

Table 3. Averaged intrajudge correlation coefficients (c.c.) and corresponding probability level, based on the two ratings (test-retest) the 27 listener judges gave on 10 speech samples.

not creaky"). The small number of three raters is already enough to present a reliable --not necessarily valid-- estimation of age.

Table 3 gives information on intrajudge reliability for the scales. Values are based on calculating a correlation coefficient between scores on scales at the beginning and at the end of the rating experiment. Only mean values are given, as the raters could be regarded as one group. Apart from the value for the scale "slovenly-polished" (0.69), which comes as an extreme compared to other figures, correlation coefficients range from 0.87 ("dull-clear") to 0.95 ("creaky-not creaky") for the bipolar semantic scales. Estimated age shows a high value of 0.99. Generally, it can be concluded that no change in using the scaling instrument occurred over the experiment as a whole.

Mean scores

Table 4 gives calculated mean values and standard deviations for the 92 untrained and 42 trained female, and 47 untrained and 43 trained male subjects on each scale. With a few exceptions, mean values range between scale values -1 and 1 with standard deviations ranging from 0.6 to 1.0. Compared to untrained females, trained ones show higher averaged values on all scales, that is, their speech is rated more positively on an averaged base. Compared to untrained males, trained males have both higher and lower averaged values. Speech of trained males is especially rated more breathy, creaky, dull and broad on an averaged base.

No debate about gender will be held here, because of the inconsistent differences in mean scale values, as well as lacking information on statistical significance of these differences (see next section for significance of differences).

Except for untrained men, mean estimated ages are within 4 years of the correct mean age, which shows, on an averaged base, the good impression we get of age by listening to speech of a person.

Group differences in mean scores

To analyze significant differences between men and women, and untrained and trained subjects ANCOVA was performed. Table 5 gives the results of these analyses. With the information on mean scale values from table 4 the following observations can be made: Regarding gender, speech of women was rated more expressive ($p<0.001$) and melodious ($p<0.001$), higher ($p<0.001$), and clearer ($p=0.032$) on the one hand, and more unpleasant ($p=0.005$), softer ($p=0.037$), and weaker ($p<0.001$) on the other.

	Scale	Women Untrained	Women Trained	Men Untrained	Men Trained
1	expressionless-expressive	0.68 (0.81)	1.13 (0.70)	0.30 (0.86)	0.58 (0.98)
2	monotonous-melodious	0.81 (0.77)	1.24 (0.67)	0.38 (0.84)	0.59 (0.93)
3	slurring-sprightly	0.41 (0.72)	0.58 (0.78)	0.24 (0.83)	0.27 (0.92)
4	shrill-warm	-0.22 (0.68)	0.36 (0.78)	1.00 (0.65)	1.03 (0.63)
5	high-low for a (wo)man	-0.57 (0.59)	-0.28 (0.59)	0.44 (0.49)	0.52 (0.60)
6	ugly-beautiful	-0.23 (0.74)	0.35 (0.81)	0.42 (0.87)	0.32 (0.93)
7	unpleasant-pleasant	-0.20 (0.77)	0.41 (0.83)	0.34 (1.00)	0.42 (1.03)
8	breathy-not breathy	0.38 (0.90)	0.87 (0.74)	0.87 (0.65)	0.69 (1.01)
9	creaky-not creaky	0.59 (0.65)	0.65 (0.77)	0.48 (0.72)	0.26 (0.80)
10	dull-clear	0.57 (0.63)	0.80 (0.64)	0.42 (0.64)	0.33 (0.80)
11	soft-loud	0.36 (0.67)	0.36 (0.64)	0.51 (0.64)	0.67 (0.58)
12	weak-powerful	0.10 (0.67)	0.31 (0.63)	0.55 (0.68)	0.64 (0.67)
13	broad-cultured	0.25 (0.87)	0.86 (0.80)	1.03 (0.60)	0.32 (0.91)
14	slovenly-polished	0.43 (0.78)	1.22 (0.69)	0.58 (0.82)	0.77 (0.70)
15	estimated age	23.72 (4.39)	32.98 (8.39)	31.37 (3.16)	44.56 (8.96)

Table 4. Mean score values and standard deviations (between brackets) of groups on the 15 scales.

	Scale	Gender		Vocal training		Gender x Training		Age		
		F	p	F	p	F	p	F	p	rrc
1	expressionless-expressive	13.79	<0.001*	5.80	0.017*	0.57	0.449	0.53	0.470	0.003
2	monotonous-melodious	18.86	<0.001*	5.64	0.018*	0.87	0.352	0.04	0.836	0.001
3	slurring-sprightly	2.68	0.103	1.79	0.183	0.23	0.630	0.60	0.438	-0.003
4	shrill-warm	101.73	<0.001*	10.30	0.002*	7.56	0.006*	30.64	<0.001*	0.018
	untrained/women resp.	100.59	<0.0001*	19.00	<0.0001*					
	trained/men resp.	18.81	<0.0001*	0.04	0.843					
5	high-low for a wo(man)	115.11	<0.001*	2.30	0.131	1.94	0.165	37.51	<0.001*	0.016
6	ugly-beautiful	13.47	<0.001*	8.58	0.004*	7.33	0.007*	2.87	0.092	0.006
	untrained/women resp.	21.65	<0.0001*	16.90	<0.001*					
	trained/men resp.	0.02	0.891	0.29	0.593					
7	unpleasant-pleasant	8.00	0.005*	8.87	0.003*	3.78	0.053	5.22	0.023*	0.009
8	breathy-not breathy	4.41	0.037*	3.60	0.059	7.17	0.008*	1.15	0.284	0.004
	untrained/women resp.	11.15	0.001*	9.51	0.003*					
	trained/men resp.	0.85	0.358	1.07	0.303					
9	creaky-not creaky	2.51	0.115	0.33	0.566	1.15	0.285	6.03	0.015*	-0.008
10	dull-clear	4.64	0.032*	4.72	0.031*	1.67	0.198	4.53	0.034*	-0.007
11	soft-loud	4.39	0.037*	0.02	0.883	0.57	0.450	3.60	0.059	0.006
12	weak-powerful	16.65	<0.001*	1.63	0.203	0.43	0.511	7.89	0.005*	0.009

Table 5. Analysis of (co)variance summary table with effects of grouping variables and covariant on scales.
df = 1 , 222 for gender, vocal training, and gender x training. rrc, raw regression coefficient. * $p < 0.05$.

Regarding vocal training, speech of trained subjects was rated more expressive (p=0.017), melodious (p=0.018) and pleasant (p=0.003), and also clearer (p=0.031).

A significant interaction between grouping variables gender and vocal training was found for the scales "shrill-warm" (p=0.006), "ugly-beautiful" (p=0.007), "breathy-not breathy" (p=0.008), "broad-cultured" (p<0.001), and "slovenly-polished" (p=0.004). Therefore, separate oneway ANOVAs were performed at each grouping variable level. Speech of untrained women was rated more shrill (p<0.0001), ugly (p<0.0001), breathy (p=0.001) and broad (p<0.0001), as compared to speech of untrained men. Compared to trained men, speech of trained women was also rated more shrill (p<0.0001), however, it was rated less broad (p=0.005) and slovenly (p=0.003). Speech of untrained women was rated more shrill (p<0.0001), ugly (p<0.0001), breathy (p=0.003), broad (p<0.001) and slovenly (p<0.0001), as compared to speech of trained women. Finally, speech of trained men was rated more broad (p<0.0001), as compared to speech of untrained men.

Age of the speaker had on the one hand a significantly positive influence on the scales "shrill-warm" (p<0.001), "high-low" (p<0.001), "unpleasant-pleasant" (p=0.023), "weak-powerful" (p=0.005) and "slovenly-polished" (p=0.003), and a significantly negative influence on the scales "creaky-not creaky" (p=0.015) and "dull-clear" (p=0.034) on the other.

Estimated age was rated significantly different for the grouping variables gender (p<0.001) and vocal training (p=0.001), men being older than women, and trained subjects being older than untrained ones. Age had a highly positive significant influence on estimated age (F(1,222)=1227.06; p<0.001).

Factor loadings

Table 6 shows the loadings, after varimax rotation, of each of the 15 scales on the six main factors extracted by means of factor analysis.

A clear factor structure emerges from the analysis. The percentage of the variance accounted for by six factors was 93.0. The eigenvalues of the six factors (i.e. the summation of the squared loadings on the 15 scales on each factor, expressing the amount of variation in the scales explained by that factor) after varimax rotation were: factor 1, 6.11; factor 2, 2.90; factor 3, 1.99; factor 4, 1.60; factor 5, 0.81; factor 6, 0.54. The communality h^2 (i.e. the summation of the squared loadings of one scale on each of the factors) ranged from 0.83 ("creaky-not creaky") to 0.97 ("expressionless-expressive" and "weak-powerful").

A first factor emerged with high loadings on the scales "expressionless-expressive", "monotonous-melodious" and "slurring-sprightly". Because all three scales are related to prosodic features of speech, this factor was given the term Intonation.

The second factor was a combined evaluative factor, with a strong component representing pitch level by the scales "shrill-warm" and "high-low for a wo(man)" on the one hand, and a qualitative component represented by the scales "ugly-beautiful" and "unpleasant-pleasant" on the other. As the scale "shrill-warm" also represents an emotional impression of speech, this factor was given the term Quality.

In the third factor high loadings were found on the scales "breathy-not breathy", "creaky-not creaky" and "dull-clear", all representing characteristics of the voice source. Therefore this factor was labelled Physiology.

A fourth factor was labelled Dynamics, because of the high loadings on the scales "soft-loud" and "weak-powerful".

	Scale	Loadings of the scale on the factor						h
		1:Intonati-on	2:Quality	3:Physiology	4:Dynamics	5:Articulation	6:Estimated age	
1	expressionless-expressi-ve	0.93	-0.05	0.10	0.19	0.22	0.03	0.97
2	monotonous-melodious	0.92	-0.10	0.13	0.18	0.23	0.00	0.96
3	slurring-sprightly	0.91	0.04	0.16	0.23	0.09	-0.11	0.93
4	shrill-warm	-0.05	0.92	0.12	0.07	0.25	0.14	0.95
5	high-low for a wo(man)	-0.33	0.86	-0.12	0.18	0.04	0.21	0.94
6	ugly-beautiful	0.39	0.64	0.46	0.12	0.40	0.01	0.95
7	unpleasant-pleasant	0.45	0.62	0.43	0.13	0.39	0.05	0.93
8	breathy-not breathy	-0.02	0.08	0.92	0.24	0.10	0.07	0.92
9	creaky-not creaky	0.17	0.19	0.79	-0.24	0.02	-0.29	0.83
10	dull-clear	0.39	-0.10	0.79	0.29	0.25	-0.08	0.94
11	soft-loud	0.25	0.06	0.07	0.93	-0.09	0.03	0.95
12	weak-powerful	0.32	0.29	0.20	0.85	-0.01	0.06	0.97
13	broad-cultured	0.21	0.28	0.03	-0.03	0.82	-0.26	0.87
14	slovenly-polished	0.27	0.19	0.24	-0.11	0.81	0.24	0.88
15	estimated age	-0.03	0.26	-0.13	0.05	-0.01	0.93	0.94
Eigenvalue		6.11	2.90	1.99	1.60	0.81	0.54	

Table 6. Factor analysis of the 14 bipolar semantic scales and estimated age scale used in the rating experiment. The rows show the scales' varimax rotated factor loadings on the six main factors, labelled Intonation, Quality, Physiology, Dynamics, Articulation and Estimated age and their communalities (h^2). The bottom rows show eigenvalues, percentage of variance accounted for, and cumulative percentage of variance accounted for by the factors. Factor loadings of 0.45 and higher are in italic.

A fifth factor was labelled Articulation, because of the high loadings on the scales "broad-cultured" and "slovenly-polished".

The scale "estimated age" was uniquely represented with a high loading on a separate factor, which therefore also was given the term Estimated age.

Group differences in factor scores

Factor scores were calculated for each of the 224 subjects. Because of the large number of subjects, no overview of individual factor scores will be given. Instead, table 7 summarizes mean values and standard deviations of the factor scores for each group. Except for factor 4, Dynamics, trained women have more positive mean values than the untrained ones, expressing the higher appreciation of their speech by the listeners. Trained men have a more positive mean value on factor 1, Intonation, while, compared to untrained men, their

	Women		Men	
Scale	Untrai-ned	Trained	Untrai-ned	Trained
Factor 1; Intonation	0.13 (0.89)	0.41 (0.86)	-0.54 (0.97)	-0.11 (1.14)
Factor 2; Quality	-0.49 (0.84)	-0.23 (1.04)	0.64 (0.84)	0.59 (0.76)
Factor 3; Physiology	-0.03 (1.01)	0.26 (0.91)	-0.04 (0.84)	-0.14 (1.19)
Factor 4; Dynamics	-0.20 (1.06)	-0.23 (0.92)	0.37 (0.93)	0.26 (0.85)
Factor 5; Articulation	-0.25 (1.00)	0.56 (0.96)	0.22 (0.82)	-0.24 (0.96)
Factor 6; Estimated age	-0.58 (0.56)	0.40 (1.03)	-0.32 (0.44)	1.19 (0.96)

Table 7. Mean factors scores and (between brackets) standard deviations.

articulation is less appreciated, regarding the more negative mean value on factor 5.

ANCOVAs were performed to analyze significant influences of grouping variables and covariant age on factor scores. Table 8 gives the results. Gender had a significant effect on factor 1, Intonation ($p<0.001$); factor 2, Quality ($p<0.001$); and factor 4, Dynamics ($p<0.001$), women having a more positively

Scale	Gender F	Gender p	Vocal training F	Vocal training p	Gender x Training F	Gender x Training p	Age F	Age p	Age rrc
Factor 1 sum score; Intonation	21.24	<0.001*	2.15	0.144	0.19	0.665	0.37	0.544	0.003
Factor 2 sum score; Quality	66.39	<0.001*	1.92	0.167	1.23	0.269	9.21	0.003*	0.012
Factor 3 sum score; Physiology	0.85	0.359	1.18	0.278	1.62	0.205	0.14	0.711	-0.002
Factor 4 sum score; Dynamics	16.03	<0.001*	0.02	0.889	0.31	0.861	0.32	0.575	0.003
Factor 5 sum score; Articulation	0.29	0.588	9.91	0.002*	18.91	<0.001*	0.47	0.492	-0.003
untrained/women resp.	7.84	0.006*	19.40	<0.0001*					
trained/men resp.	14.58	<0.001*	5.98	0.017*					
Factor 6 sum score; Estimated age	1.59	0.209	9.12	0.003*	0.48	0.489	596.82	<0.001*	0.059

Table 8. Analysis of (co)variance summary table with effects of grouping variables and covariant on factor sum scores.
df = 1, 222 for gender, vocal training, and gender x training. rrc, raw regression coefficient. * p < 0.05

rated Intonation characteristic on the one hand, and more negatively rated Quality and Dynamics characteristics on the other.

A significant interaction between grouping variables was found for factor 5, Articulation (p<0.001). Therefore, separate oneway ANOVAs were performed at each grouping variable level. Untrained women have a significantly more negatively rated Articulation (p=0.006) compared with untrained men, while the opposite was observed in trained subjects (p<0.001). Untrained women have a significantly more negatively rated Articulation (p<0.0001) compared with trained women, while Articulation characteristic of speech of untrained men is rated significantly more positive (p=0.017), as compared to trained men.

Although significant influences of vocal training were found on scales, no specific influences were observed on calculated factor scores. Estimated age being the only scale present in factor 6, again showed to be significantly influenced by vocal training (p=0.003), untrained subjects having a younger rated age.

Age as a covariant had a significantly positive influence on factor 2, Quality, implying that speech of older subjects is more appreciated.

Discriminant analysis

Discriminant analysis was performed to analyze which scales best differentiate between male and female speech, and, separately for men and women, between untrained and trained speech. The statistical procedure was separately performed for men and women, because of the significant interactions between grouping variables Gender and vocal training on several scales (see table 5). Estimated age was not introduced in the analysis, because it is not a bipolar semantic scale.

The discriminant analysis on the scale means per speaker determined the scales "weak-powerful", "monotonous-melodious", "shrill-warm", "creaky-not creaky", "slurring-sprightly", "unpleasant-pleasant", "breathy-not breathy", "dull-clear", "broad-cultured" and "high-low for a (wo)man", in order of importance, to be the set of scales discriminating best between male and female speakers. The canonical correlation of the discriminant function was 0.78, which means that $(0.78)^2 \times 100 = 61\%$ of the variation in the discriminant function is explained by the groups. The percentage correctly classified in their own group on the basis of the classification function coefficients was 90.6%. Sixteen women and five men were incorrectly classified. The centroids (group means) of the two groups on the canonical discriminant function were -1.00 for women and 1.52 for men.

For the female subgroup the discriminant analysis on the scale means per speaker determined the scales "slovenly-polished", "shrill-warm", "creaky-not creaky", "breathy-not breathy", "dull-clear", "monotonous-melodious", "unpleasant-pleasant", "broad-cultured" and "slurring-sprightly", in order of importance, to be the set of scales discriminating best between untrained and trained speakers. The canonical correlation of the discriminant function was 0.63, which means that $(0.63)^2 \times 100 = 40\%$ of the variation in the discriminant function is explained by the groups. The percentage correctly classified in their own group on the basis of the classification function coefficients was 80.7%. Nineteen untrained and seven trained women were incorrectly classified. The centroids (group means) of the two groups on the canonical discriminant function were -0.54 for untrained and 1.20 for trained women.

For the male subgroup the discriminant analysis on the scale means per speaker determined the scales "broad-cultured", "slovenly-polished", "expressionless-expressive", "slurring-sprightly", "high-low", "ugly-beautiful", "unpleasant-pleasant", "dull-clear" and "shrill-warm", in order of importance, to be the set of scales discriminating best between untrained and trained speakers. The canonical correlation of the discriminant function was 0.69, which means that $(0.69)^2 \times 100 = 48\%$ of the variation in the discriminant function is explained by the groups. The percentage correctly classified in their own group on the basis of the classification function coefficients was 80.9%.

Seven untrained and nine trained men were incorrectly classified. The centroids (group means) of the two groups on the canonical discriminant function were -0.91 for untrained and 0.98 for trained men.

DISCUSSION

The results of the conducted experiment show the possibilities of evaluating speech of groups of subjects. Speech of trained subjects was used to give direction to what might be regarded as more "ideal". Untrained subjects without vocal complaints or visually observed abnormalities of the vocal folds produced the speech that was used to establish a frame of reference, consisting of averaged scores on the utilized scales. This frame of reference, which can be acknowledged as representing "normal" speech, is needed to have an image of a "normal" arrangement of perceptual dimensions. Evaluation of speech can be performed by comparing perceptual dimensions with the "normal" arrangement and to specify deviations. To facilitate this evaluation and to offer material for new investigations, speech samples used in this experiment, as well as perceptual specifications and demographic descriptions of each subject were made available on CD-ROMs (SPEX).

Scale properties

The scaling instrument used in this study to perceptually evaluate speech of groups of subjects, showed its practicability and gave proof of its careful construction. Listeners used scale ends with a same representation of a perceptual dimension. Reliably scoring of the scales required a limited number of judges (< 20). Only the scales "creaky-not creaky", "dull-clear" and "slovenly-polished" can be considered as exceptions. The first of these scales "creaky-not creaky" was introduced in this study as a new scale, because of the potential influence of this modality on the perceptive quality of speech. Creaky voice is present in normal speech and regarded as a separate mode of phonation. However, the use of this mode of phonation might differ between groups and, thus, have an influence on --overall-- ratings of speech quality. Although factor analysis of the scores on this scale showed that listeners made judgments as one group, the eigenvalue was lower compared to the other scales, indicating the difficulty that some listeners might have had in giving concise judgments. The same problem might have been present while giving ratings on the other two scales with less effective --however, still high enough-- reliabilities, "dull-clear" and "slovenly-polished". Fagel et al. (8) found changing opinions in judgments on the scale "broad-cultured". It could be that listeners differ in their tolerance regarding an other aspect of articulation as expressed in the scale "slovenly-polished". Compared to previous investigations (9,12,18) no problems were experienced while dealing with the scale "breathy-not breathy". A high effective reliability of 0.937 was found and factor analysis of the scores on this scale yielded an eigenvalue of 11.5, which explained 42.6% of the variance in the scores. The almost exclusive presence of female listeners might explain the better characteristics of the scale "breathy-not breathy" in this study, as women are known to be more associative raters with higher correlations between scales (18). The higher number of subjects incorporated in this experiment might also have presented a larger variety of breathiness in the stimulus material, thus producing a higher reliability.

Intrajudge reliability was sufficiently high, considering the high values for correlation coefficients (c.c.>0.90) with low probabilities (p<0.001). The only exception was presented by the scale "slovenly-polished" (c.c. 0.69, p=0.028). It, again, reflects the potential difficulty in rating this articulatory characteristic of speech.

The scale estimated age showed a very high inter- and intrajudge reliability, and there was a high level of agreement among listeners about the use of this scale, considering the eigenvalue of 22.1 of the first factor, explaining 81.9% of the variation in the data. Previous work already established the ability of listeners to adequately estimate age of speakers (9,19-21).

The preparation of the scaling instrument with aligning polarities probably resulted in a more practicable instrument, as listeners can more easily express their qualitative impression of aspects of speech on one side of the rating form, without having to check the correct direction of the polarity. With this aligning an improved version of the original scaling instrument is given for evaluation of laryngeal speech.

Group differences in mean scores

Many differences in scale ratings were found between men and women, as well as between untrained and trained subjects. Speech of women is characterized by more positively judged intonation features, having higher ratings on the scales "expressionless-expressive" and "monotonous-melodious". Intonation is determined for an important part by regulation and variation of the fundamental frequency (F_0) (22). Slow variation of F_0 during an utterance is especially controlled by subglottal pressure, while variation of intralaryngeal muscular activity provide local F_0 movements (22). With these considerations, women should show more variation in activity of intralaryngeal musculature during speech, compared to men. Another positive aspect of female speech is the clearer impression listeners get, which is probably caused by the acoustic characteristics of the smaller dimensions of the female vocal tract (23) and the higher F_0 of women (9).

Although the suffix "for a (wo)man" was especially added to the scale "high-low" to compensate for gender-specific differences in F_0, women were still given a significantly more negative rating on the scale "high-low". It seems that the listeners in this study were not able to compare pitch of the speaker with a gender-neutral image in this specific perceptual dimension and that pitch of men and women was rated systematically to low and high, respectively. The scale "shrill-warm" is closely related to the scale "high-low" and therefore it was no surprise that women were also rated significantly more shrill. Kreiman et al., (5) found a specific perceptual relation between rated degree of pathology and F_0. The observed differences in ratings on the scales "high-low" and "shrill-warm" could thus have an influence on judgments on the scales "ugly-beautiful" and "unpleasant-pleasant" with women having more negatively rated speech.

Compared to speech of men, speech of women was rated softer and weaker, which is in agreement with Awan (24), who measured intensity level of conversational speech of men and women and found significantly softer intensities for female speakers. A second explanation for this difference might be found in the so-called Frequency Code, which suggests that listeners perceive female speech as "small" (25). Louder and more powerful male speech might probably result in a higher intelligibility. However, women might compensate this by a more careful and correct pronunciation (26-28). In the present study only speech of trained women was rated more polished than the male counterpart, which does not give hard evidence for a higher appreciated articulation in females by the group of listeners.

Breathiness is inversely related to glottal closure (29). Speech of untrained females was rated more breathy, which is, therefore, in concordance with previously published results from our research project showing the relatively higher leakage of air (10), as well as the smaller percentage of vocal fold closure in women (3), as compared to men.

Profession and education are known to have an influence on articulation ratings (9). Ratings on the scale "broad-cultured" might, therefore, be influenced by the social background of the subjects. Nearly all untrained male and female subjects were university students or receiving vocational training, respectively, while trained subjects were recruited from choirs with a more diverse social stratification.

In her study, Tielen (9) used an almost identical scaling instrument to compare speech of untrained men and women. Our results are in general in agreement with her study, however a few differences are apparent. In the present study the male speakers are the louder and more powerful ones and women were rated more breathy. Regarding tempo of speech (scales "dragging-brisk" and "slow-quick" in the Tielen study), a same tendency was found with more positively ratings on the scale "slurring-sprightly" in women. In the Tielen study the effect of an interaction between gender and profession of speaker on the scales "ugly-beautiful" and "unpleasant-pleasant" might have prevented showing difference between men and women in these scales, a difference that can be found in the present study.

Trained speech was rated more expressive and melodious. Phonatory motor control (11) in the trained groups may have provided the subjects in these groups with better intonation abilities. Trained subjects do also have larger intensity ranges (1), which they might employ more fully during speech. The clearer speech of trained subjects might be based on differences in frequency spectrum of the voice source. Compared to untrained subjects typical spectra of singers have a relatively smaller decay in intensity level with higher harmonics and show clustering of formants, producing a so-called singers formant (30). Spectra with more information in the higher frequency region are perceptually characterized as less breathy and more sonorous (31). The changed aspect of trained spectra is related to an increased glottal closure (31). Though not statistically significant, trained subjects in our study had a higher percentage glottal closure than untrained ones (2).

Age was related to a number of scales. Older subjects have speech that is rated warmer on the one hand and more dull on the other. Both characteristics probably depend on ageing of vocal fold structures (2). The more positively rated pleasantness with age in this study is in contradiction with the finding of Tielen (9). A cause might be selection bias, as the average age of trained subjects was older than that of the untrained ones, and speech of trained subjects was rated more positively. Older subjects also received a higher rating on creaky voice, which is often associated with senescence.

Listeners were able to give a good estimation of the age of speakers. Studies showed that estimation might be based on pitch information (32-34) and reading performances (19,34,35).

Factor loadings

Factor analysis resulted in a solution with six factors labelled Intonation, Quality, Physiology, Dynamics, Articulation and Estimated age. Together they explained 93% of the variation in the scores on the scales. The most important factor in evaluating speech was Intonation. Listeners seem to have an attentive ear regarding the perception of variation of speaking fundamental frequency and speaking intensity level (scales "monotonous-melodious" and "expressionless-expressive"), as well as the ability to register the speed of variation of these variables (scales "expressionless-expressive" and "slurring-sprightly"). Positive ratings on these scales are correlated with more pleasant rated speech. Even more important for speech to be rated beautiful and pleasant are a warm and relatively low voice. Scales representing these perceptual dimensions are clustered in the second factor, Quality. The scale "ugly-beautiful" is also related to three other scales clustered in the third factor Physiology, that is, clear speech without breathiness and creaky voice is rated more beautiful. The fifth factor, Articulation, is the first one with an eigenvalue less than 1. This threshold is normally used to designate the number of factors in a solution. However, because the study of Fagel et al. (8) showed a solution with five factors, each representing a specific perceptual dimension, this threshold was not used in the present study and statistical analysis was forced to produce six factors, one of these separately designated for estimated age. Table 6 shows that cultured and polished articulation is also associated with speech that is rated as being beautiful and pleasant. Estimated age is associated with polished, though broad, speech with a creaky and low voice (see Table 6).

With their data on laryngeal and alaryngeal speech Fagel et al. (8), Nieboer et al. (12) and Tielen (9) followed the same design in evaluating perceptual scores. Typically constructed for the evaluation of alaryngeal speech, Nieboer et al. (12) introduced new scales and left out others, resulting in a factorial solution that is hard to compare with the solution of the present study. In the other two studies comparable factorial solutions emerged. However, labelling of the factors among the studies varied, due to minor differences in scale composition and relative contribution of scales to the specific factors.

Group differences in factor scores

Table 7 presents data on perception of speech in a more comprehensive way by giving means of factor scores for each group. Significance of differences in scores between groups are given in Table 8. The grouping variable gender has a significant effect on Intonation, Quality, Dynamics and Articulation, expressing the large differences that can be obtained while perceptually evaluating speech of men and women. Although differences were found on scale level when comparing speech of trained and untrained subjects, differences were less explicit using factor scores. A significant effect of the grouping variable vocal training was found only on Articulation.

An encouraging aspect of ageing is the higher appreciation listeners have for speech of older persons, regarding the positive relation between age and Quality.

Discriminant analysis

A high percentage of 91% of the subjects were correctly classified for gender. Scales used for this classification come from several factors. The most important scale is "weak-powerful", which refers to the Frequency Code of Ohala (25), suggesting that "size" of male speech is perceived as "large", due to its lower pitch. The second scale in the discriminant function is "monotonous-melodious", referring to the more emotional impression that listeners have of female speech (9). Scales from the factor Physiology are also important for a correct classification: "creaky-not creaky", "breathy-not breathy" and "dull-clear" have a place in the discriminant function. It points to the difference in vocal function between men and women (10).

In the female subgroup 81% of the subjects were correctly classified for trained or untrained status. Articulation and intonation are processes that can be regulated actively for an important part. These processes, categorized as two separate factors in the present study, are perceptively represented by the scales "slovenly-polished", "monotonous-melodious", "broad-cultured" and "slurring-sprightly" in the discriminant function. The more positive ratings that were given to speech of trained women and the use of these scales for a correct classification stress the presumably more precise articulation, as well as the higher appreciated variation of the variables pitch and intensity in trained women. These qualities of trained women might be based on the experienced level of motor control over the vocal and articulatory organs (1,11). However, a biasing influence of age might be present, considering the positive relation between age and the scale "slovenly-polished". A process merely beyond active control concerns glottal function. A correct classification is also based on specific ratings on the scales clustered in the factor Physiology, trained women showing the more positive ratings. It suggests a difference in vocal fold function between trained and untrained women; however, a related study concentrating on physiology of phonation did not show clear differences between untrained and trained women (10).

A correct classification of 81% was also found for vocal training in men. Important scales for this classification are clustered in the factors Intonation and Articulation, which might both be positively influenced by vocal training, as discussed in the previous paragraph. Physiology seems to be of less importance for a correct classification in the male subgroup.

CONCLUSIONS

With the information given in the previous sections the questions given in the introduction can be answered.

1. Speech of a large number of voice healthy subjects can reliably be evaluated with a scaling instrument such as suggested by Fagel et al. (8).

2. Factor analysis resulted in a solution with six factors, labelled Intonation, Quality, Physiology, Dynamics, Articulation and Estimated age, which represent diverse aspects of speech. With these factors further evaluation of perceptive data on speech can be expedited.

3. Many significant differences in the perception of speech of men and women were found. Speech of women was rated more expressive, melodious, breathy and shrill, higher and clearer on the one hand, and more unpleasant, softer and weaker on the other. Regarding vocal training, speech of trained subjects was rated more expressive, melodious, and pleasant, and also clearer.

4. Significant differences between men and women were also found on factor level. Intonation of women was judged more positively by listeners, whereas more positive ratings were given on Quality and Dynamics of speech of men. On factor level no significant differences were found between speakers with and without vocal training. Social background and education level of subjects demonstrated to have an influence on Articulation.

5. With discriminant analysis scales were selected that can best be used to classify subjects in male and female, and trained and untrained groups. For classification of gender scales were selected from the factors Dynamics ("weak-powerful"), Intonation ("monotonous-melodious" and "slurring-sprightly"), Quality ("shrill-warm", "unpleasant-pleasant" and "high-low for a [wo]man") and Physiology ("creaky-not creaky", "breathy-not breathy" and "dull-clear"). Regarding classification for vocal training, in the female subgroup scales were selected from the factors Articulation ("slovenly-polished" and "broad-cultured"), Quality ("shrill-warm" and "unpleasant-pleasant"), Physiology ("creaky-not creaky", "breathy-not breathy" and "dull-clear") and Intonation ("monotonous-melodious" and "slurring-sprightly"), whereas in the male subgroup scales were selected from the factors Articulation ("broad-cultured" and "slovenly-polished"), Intonation ("expressionless-expressive" and "slurring-sprightly") and Quality ("high-low", "ugly-beautiful", "unpleasant-pleasant" and "shrill-warm").

References

1. Sulter A, Schutte H, Miller D. Differences in phonetogram features between male and female subjects with and without vocal training. J Voice 1995;9:363-77.
2. Sulter A, Schutte H, Miller D. Standardized laryngeal videostroboscopic rating: Differences between untrained and trained male and female subjects, and effects of varying sound intensity, fundamental frequency and age. J Voice 1996;10:175-89.
3. Sulter A, Albers F. (in press). The effects of frequency and intensity level on glottal closure in normal subjects. Clin Otolaryngol 1996:in press.
4. Fex S. Perceptual evaluation. J Voice 1992;6:155-8.
5. Kreiman J, Gerratt B, Precoda K. Listener experience and perception of voice quality. J Speech Hear Res 1990;33:103-15.
6. Kreiman J, Gerratt B, Precoda K, Berke G. Individual differences in voice quality perception. J Speech Hear Res 1992;35:512-20.
7. Kuwabara H, Ohgushi K. Experiments on voice qualities of vowels in males and females and correlation with acoustic features. Lang Speech 1984;27:135-45.
8. Fagel W, Van Herpt L, Boves L. Analysis of the perceptual qualities of Dutch speakers' voice and

pronunciation. Speech Comm 1983;2:315-26.
9. Tielen M. Male and female speech. An experimental study of sex-related voice and pronunciation characteristics . Thesis, Universiteit van Amsterdam, 1992.
10. Sulter A, Wit H. Glottal volume velocity waveform characteristics in subjects with and without vocal training, and relations with gender, sound intensity, fundamental frequency and age. J Acoust Soc Am 1996:in press.
11. Murray T, Caligiuri M. Phonatory and nonphonatory motor control in singers. J Voice 1989;3:257-63.
12. Nieboer G, De Graaf T, Schutte H. Esophageal voice quality judgements by means of the semantic differential. J Phonetics 1988;16:417-36.
13. Osgood C, Suci G, Tannenbaum P. The measurement of meaning . Urbana, IL: University of Illinois Press, 1957.
14. Kim J. Factor analysis. In: Nie H, et al. eds. SPSS: Statistical package for the social sciences . New York: McGraw-Hill Book Company, 1975:468-514.
15. Hatch E, Farhady H. Research design and statistics for applied linguistics . Rowley, London , Tokyo: Newbury House Publishers, 1982.
16. SPSS Inc. SPSSX user's guide (2nd ed.). New York: McGraw-Hill Book Company, 1986.
17. Nunnally J. Psychometric theory (2nd ed.). New York: McGraw-Hill Book Company, 1978.
18. Van Herpt L. Do men and women use a common semantic factor space to describe voice and pronunciation? Proceedings of the Institute of Phonetic Sciences, University of Amsterdam 1987;11:15-25.
19. Ptacek P, Sander E. Age recognition from voice. J Speech Hear Res 1966;9:273-7.
20. Shipp T, Hollien H. Perception of the aging male voice. J Speech Hear Res 1969;12:704-10.
21. Hartman D, Danhauer J. Perceptual features of speech for males in four perceived age decades. J Acoust Soc Am 1976;55:713-5.
22. Strik H, Boves L. A physiological model of intonation. Proceedings Dept. of Language and Speech, University of Nijmegen 1992/1993;16/17:96-105.
23. Fant G. Non-uniform vowel normalization. STL-QPSR 1975;2-3:1-19.
24. Awan S. Superimposition of speaking voice characteristics and phonetograms in untrained and trained vocal groups. J Voice 1993;7:30-7.
25. Ohala J. Cross-language use of pitch: an ethological view. Phonetica 1983;40:1-18.
26. Smith P. Sex markers in speech. In: Scherer K, Giles H, eds. Social markers in speech 1979:109-46.
27. Koopmans-van Beinum F. Vowel contrast reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions . Thesis. Universiteit van Amsterdam, 1980.
28. Henton C. A comparative study of phonetic sex-specific differences across languages . Thesis, Oxford, Great Britain, 1985.
29. Södersten M, Lindestad P. Glottal closure and perceived breathiness during phonation in normally speaking subjects. J Speech Hear Res 1990;33:601-11.
30. Sundberg J. The acoustics of the singing voice. Sci Am 1977;236:82-91.
31. Södersten M, Hammarberg B. Effects of voice training in normal-speaking women : Videostroboscopic, perceptual, and acoustic characteristics. Scand J Log Phoniatr 1993;18:33-42.
32. Hollien H, Shipp T. Speaking fundamental frequency and chronological age in males. J Speech Hear Res 1972;15:155-9.
33. Jacques R, Rastatter M. Recognition of speaker age from selected acoustic features as perceived by normal young and older listeners. Folia Phoniatr 1990;42:118-24.
34. Shipp T, Qi Y, Huntley R, Hollien H. Acoustic and temporal correlates of perceived age. J Voice 1992;6:211-6.
35. Brown Jr. W, Morris R, Michel J. Vocal jitter in young adult and aged female voices. J Voice 1989;3:113-9.

7

A structured approach to voice range profile (phonetogram) analysis

Arend M. Sulter, Hero P. Wit, Harm K. Schutte* and Donald G. Miller*

Department of Otorhinolaryngology, University Hospital Groningen and *Voice Research Lab, Department of Medical Physiology, Faculty of Medicine, University of Groningen, Groningen, the Netherlands

Previously published in the Journal of Speech and Hearing Research 1994,37,1076-1085.

INTRODUCTION

Phonetography is a practicable and readily accessible method to investigate and map the quantitative potentialities of vocal output (1-4). The maximally loud and soft phonations throughout the entire frequency range are indicated in a plot of frequency against sound pressure level (SPL).

Figure 1 gives an example of a normal phonetogram[1] from a male subject without vocal complaints. The datapoints in the plot are acquired during a short session in which the investigator asks the subject to phonate as loudly and softly as possible at selected frequencies,[2] thereby covering the subject's whole

[1]Many synonyms of the graphical representation of an individual's voice potentialities are proposed in articles concerning phonetography. The terms phonetogram, phonogram, voice range profile, voice field, voice area and F_0-SPL profile have been used in literature. In this article the term phonetogram is used.

[2]The frequency range was sampled at four pitches per octave, e.g. C3-E3-G3-A3 (in this octave: 131, 165, 196, 220 Hz).

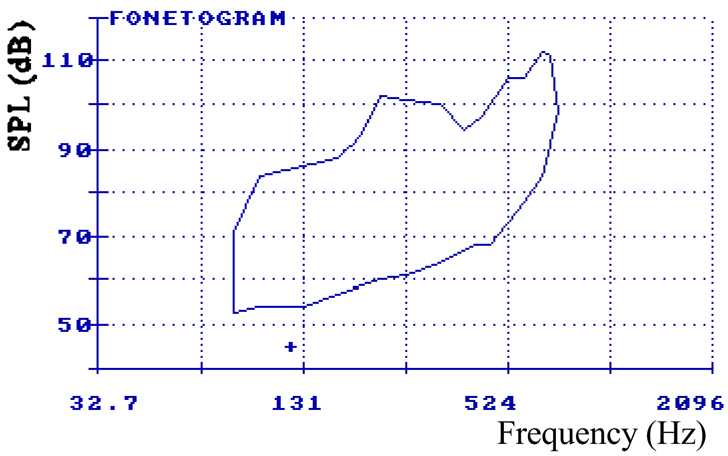


Figure 1. An example of a "normal" male phonetogram. Along the x-axis the frequency scale is plotted (32.7 - 2096 Hz) and the intensity level is given along the y-axis (40 - 120 dB). The "+" sign at 123 Hz indicates the mean speaking fundamental frequency. Note the dip in the loud phonation contour at about 400 Hz. This local minimum exhibits the transition of chest register to falsetto register.

frequency range (4,5).

The basic instrumentation consists of a tone generator producing a vowel-like sound that is used as a pitch target, and an SPL measuring device (6). The fundamental instrumentation has not changed over the years; however, the use of modern electronics considerably facilitates the operation of both instrumental components. The tone generator and the SPL measuring device have been incorporated in a computer (7), reducing the time required for both the acquisition of data and the graphical conversion into a phonetogram. This makes a visual feedback of the measurements possible for both subject and investigator. A further contribution to the automatic registration of phonetograms has been the incorporation of a unit into the equipment to determine fundamental frequency (7-9). The benefit of this unit is twofold: subjects or patients not able to sustain the given pitch can use an alternative (freely chosen) pitch. In addition, the occurrence of octave-errors and other mistakes in determining the correct pitch (which are already small when the registration is performed by experienced investigators) will be minimal. The computer also makes it possible to create immediately processable phonetographic data files.

After the first description of phonetogram-like profiles by Wolf, Stanley, & Sette (10) and an early article by Calvet & Malhiac (1), the method received sporadic attention in the literature (4,5,11-14). In recent years, however, a growing number of practical and theoretical articles on vocal function and voice use have dealt with phonetography. Recommendations were formulated to standardize procedures in the acquisition of phonetograms (6,15); the potential of phonetography as a clinical tool was illustrated (3,16,17); and the theoretical bases of profiles were questioned (2,18-20).

The practical uses of phonetography, as reflected in the literature, can be summarized as: (a) assessing information about individual voice potentialities, (b) investigating the influence of therapy or surgical intervention and (c) comparing phonetograms of selected groups (11,21-23).

The lack of clear parameters applying to the phonetogram as a whole, however, presents an obstacle in the comparison of one phonetogram with another, as well as in the establishment of standard reference values for the phonetogram.

Approaches in dealing with this problem are based on averaging (10), or rescaling techniques. With the latter technique the individual phonetogram is rescaled with the x-axis (frequency range) to 100% (3,11,21,23). After a number of phonetograms have been normalized in this way, summary statistics on intensities of vocal output can be compiled. Frequency-dependent intensity information, however, cannot be derived from these statistics.

In another approach Klingholz and Martin (2) have attempted to describe mathematically (half axis, vertices and rotation) an arbitrary number of ellipses that can be fitted on to phonetograms. However, the number of ellipses contained in a phonetogram is not specified, and there are various ways of fitting an ellipse through datapoints. Also, the acquisition of datapoints introduces an unpredictable deviation from the ideal ellipse shape. This lack of a consistent basis for analyzing the phonetogram with ellipses calls into question the validity of the results.

A different approach toward the analysis of individual phonetograms is proposed in this research note. Parameters representing three expert-acknowledged features are extracted from phonetograms. Advantages of this method include (a) the derivation of features from phonetograms without distorting its shape, and (b) the particular attention paid to the dynamic possibilities of the F_0-SPL range used in normal speech. To demonstrate this method of automated evaluation, a normal male phonetogram as well as a pathologic male phonetogram are processed, and the resulting parameter values are compared with normative male data. A future article will present these normative data and data of groups of subjects that have received vocal training over a period of at least two years (24).

METHODS

Features of phonetograms

A group of four speech therapists and three Ear, Nose, and Throat (ENT)-physicians were informally asked to describe the way they visually analyzed phonetograms and to give their opinion about what features should be regarded as important characteristics. The descriptions offered by this group included three common features:

Shape. The experts considered the shape a very important feature. The general shape of the phonetogram is complex, but it can be seen as the sum of two overlapping ellipses, each with a different slope of the long axis (2). The intersection where the two ellipses meet in the loudest phonation contour is a typical characteristic of the phonetogram of subjects without voice training. In that specific place, in male subjects at about G4 (392 Hz) and in phonetograms of women slightly higher, at about A4 (440 Hz),[3] a local minimum can be seen (see Figure 1). This local dip can be attributed to the transition from chest to falsetto register when the phonetographic datapoints are measured for the vowel /a/ (18). This interruption in the otherwise rising contour of maximum SPL is minimized by vocal training (5,11).

Enclosed area. Connection of the lines of the loud phonation contour with the soft phonation contour (the upper and lower part of the phonetogram respectively) yields an enclosed area. All observations and judgments of phonetograms take this area into account. However, lack of quantitative knowledge about what constitutes a "normal" area results in a qualitative judgment with an imaginary frame of reference. The same can be said about the frequency range: a minimum of two octaves is often used in practice (3,5,25). However, only limited knowledge is available concerning the mean range and standard deviation of the frequency range in large specified groups of men and women.

"Speaking Range" dynamics While the phonetogram covers the entire frequency range, the speaking voice in its normal function uses only a part of the range. In order to reflect the importance of this portion of the range, a formula was devised to analyze it with respect to mean speaking fundamental frequency (mff).

Parameters describing the features

Representative parameters can be defined to describe in an approximate way the different features (shape, enclosed area and "speaking range" dynamics). A relatively large number of parameters (for the feature shape) are introduced in order to increase the chance of detecting deviations from a normal pattern. A second purpose is to promote the emergence of constellations of parameter values specific to pathologic entities. Because of the large number of parameters, however, considerable redundancy can be expected.

---

[3]The x-axis of a phonetogram extends from C_1 to c^5 in the Helmholtz notation or from C1 (32.7 Hz) to C8 (2096 Hz) in American notation (6).

Shape-related parameters. Fourier Descriptors. Rather than normalizing the frequency range or mathematically describing arbitrarily projected ellipses in the phonetogram, the shape itself is analyzed with Fourier Descriptors (FD). Fourier Descriptors were developed for computerized reading of handwritten alphanumeric characters (26). In this procedure a closed contour, consisting of straight line segments, is transformed into a set of slope values as a function of length along the contour. Starting the analysis at a certain point on the perimeter (in our case the point corresponding to the lowest loud phonation frequency), one proceeds clockwise in steps (see Figure 2a). Each step consists of the computation of the length of a line segment and the angle between this segment and the following one. This procedure results in a set of length values and a set of angle values, giving angle as a discrete function of length along the contour (see Figure 2b). Formulas were developed by Zahn and Roskies (26) to calculate the Fourier transform of this function, taking into account that the points along the length-axis of the function (taken as the independent variable) are in general not equally spaced.

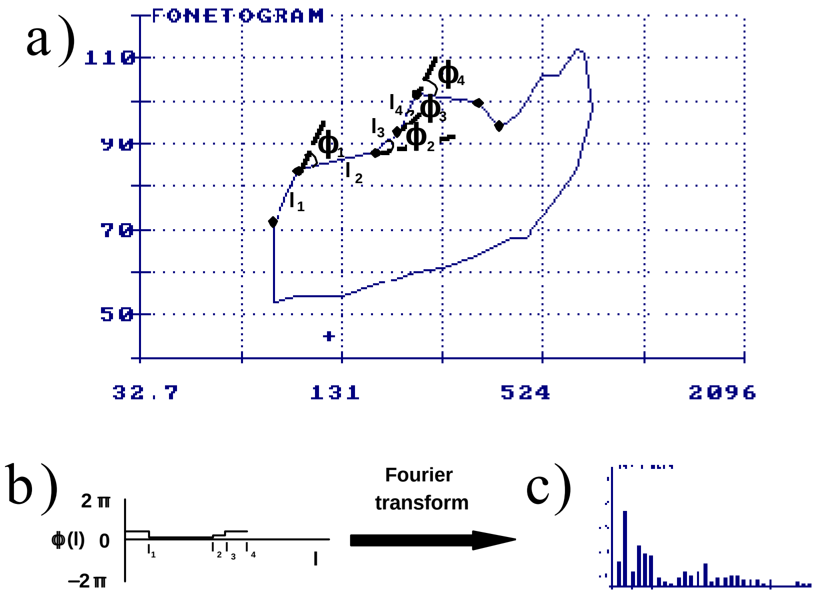


Figure 2. The shape of the phonetogram is analyzed with Fourier Descriptors. a) Starting at the lowest loud phonation, lines are drawn between the phonetogram points in a clockwise direction (l_1, l_2,..l_n). Next, the angle between adjacent lines is calculated (N_1, N_2,..N_n). b) Line lengths and angles are placed in a plot with new axes. c) The information in the plot with the length of line segments and angle axes is processed with a Fourier transform, resulting in a number of Fourier Descriptors. Close to the origin the general shape is defined, whereas the Fourier Descriptors higher on the x-axis represent small changes in shape. The amplitude of an FD gives its relative contribution to the shape.

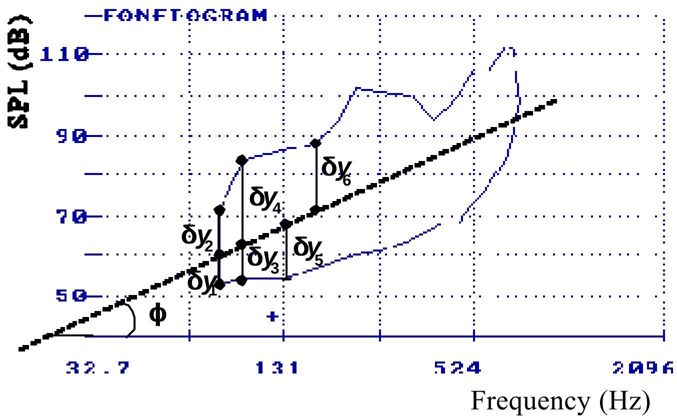


Figure 3. With a least square fit a line can be drawn through the phonetogram. This line has a minimal distance to all phonetogram points (E*y_1,..*y_n). With the defined line the angle with the x-axis can be calculated (N).

As a first attempt to investigate the usefulness of this shape quantification procedure, amplitude values of the calculated Fourier transform are displayed. These amplitude values are called "Fourier Descriptors". In the plot of Figure 2c on the x-axis a discrete number of thirty Fourier Descriptors are given. The lowest numbers define the general shape, whereas the FDs higher on the x-axis represent small changes in shape. The y-axis gives the amplitude of each FD representing its relative contribution to the shape. As an example one can consider the FDs for a circle and an ellipse: For the sake of simplicity, the length intervals with which the contour of the circle or ellipse is sampled are assumed to be equal. In the case of a circle, angle as a function of length is constant, resulting in a value of zero for all FDs. Following the contour of an ellipse, angle as a function of length will have two maxima (for the "sharp ends" of the ellipse) and two minima (the long sides of the ellipse). The magnitude of this function gives the value of FD_2. In general FD_2 is a measure for the ellipticity of a contour. When the angle function shows three maxima and minima in tracing the contour, this will be reflected in the magnitude of FD_3, and so forth.

Contour regularity. Even when care is taken for a proper acquisition of phonetogram points by following UEP procedures (6), in many cases the perimeter (especially along the loud phonation contour) has an irregular aspect (20). The parameter which illuminates this aspect is the contour regularity. This ratio is derived by dividing the enclosed area of the phonetogram by the squared perimeter, yielding a dimensionless figure. The highest contour regularity value will be derived from a circle, with greater irregularities yielding smaller values. Deviations from the circular shape correspond to smaller contour regularity values.

Phonetogram slope. A central straight line is drawn through the phonetogram. The slope and position of the line are determined by the least possible sum of squared distances to the measured points. Figure 3 shows the procedure determining the position and slope of the central line through the phonetogram.

Area-related parameters. Enclosed area . The phonetogram is separated in a loud phonation contour (upper part) and a soft phonation contour (lower part). In case there is only one measuring point at either the lowest or the highest produced frequency, this point is regarded as belonging to both the soft and loud contour. After computing the area between the lower contour and frequency axis, this area is subtracted from the area between the higher contour and the frequency axis (see Figure 4a and 4b). The remaining enclosed area is divided by a constant, namely, the area of the rectangle with corners 32.7 Hz - 40 dB, 32.7 Hz - 110 dB, 2096 Hz - 110 dB and 2096 Hz - 40 dB (see Figure 4c). The reference area is based on axes proposed in the recommendations of the UEP (6). We chose a rectangle with the y-axis from 40 to 110 dB because in clinical practice measured vocal loudness hardly exceeds the intensity level of 110 dB, and this range has a center intensity of 75 dB. This reference line is hereafter employed in the analysis of the "Speaking Range" dynamics. By using a relative quotient, this quotient is thus independent of the scaling of x- and y-axes. Furthermore, this dimensionless ratio was chosen to increase comprehensibility: The size of determined enclosed area can be directly related to the frame of the phonetogram described above. For instance, a value of 0.238, as in figure 4, indicates that almost a quarter of the reference rectangle is covered with the

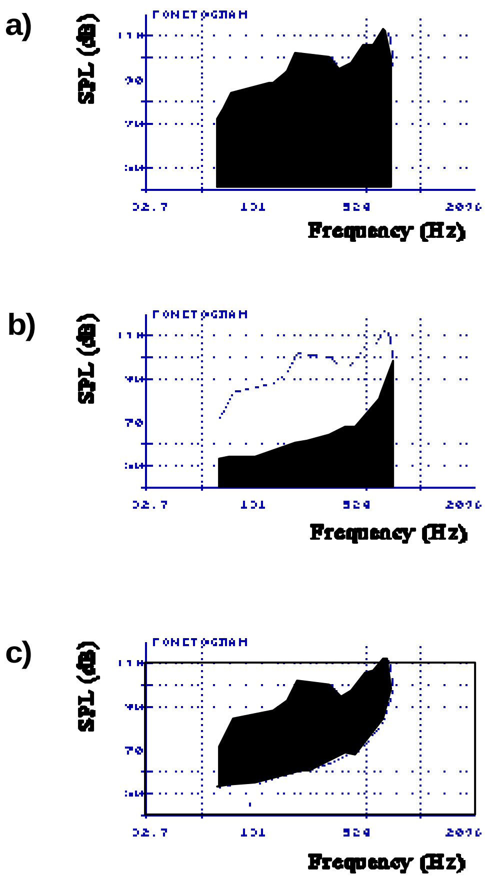


Figure 4. The enclosed area is calculated in four steps. a) The integral of the loud phonation contour is calculated. b) The integral of the soft phonation contour is determined. c) A subtraction gives the enclosed area. This area is divided by the area of the outlined rectangle. This yields a dimensionless figure relating to the part of the phonetogram covered by a subject's phonatory capabilities.

phonetogram area.

Frequency range .

$$x_{oct} = \frac{6 \times (\log (x_{Hz\ high}) - \log (x_{Hz\ low}))}{\log 2096 - \log 32.7} \qquad (1)$$

With equation (1) the individual frequency range can be obtained as a number of octaves (x_{oct}) after the highest ($x_{Hz\ high}$) and lowest ($x_{Hz\ low}$) possible phonatory frequencies have been determined and the difference between these frequencies is calculated. Figure 5 illustrates this procedure.

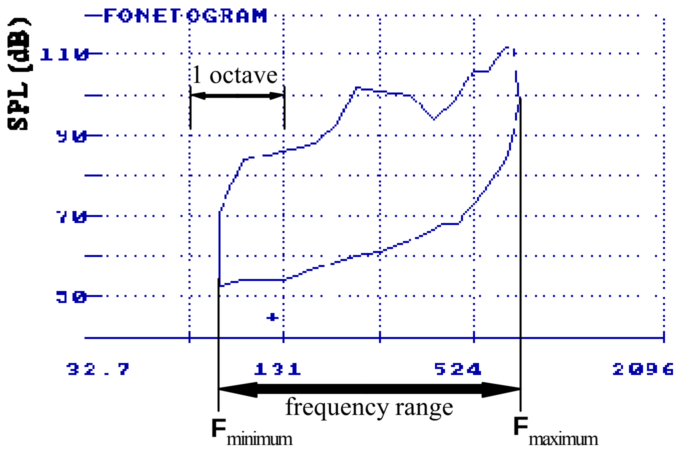


Figure 5. The lowest ($F_{minimum}$) and highest phonation ($F_{maximum}$) are transposed on a tone scale. The difference between the extremes gives the frequency range in number of octaves.

Parameters related to "Speaking Range" dynamics. At four selected frequencies the dynamic range and the central position of this range are determined. These data provide information about the capacities for a person to modulate frequency and intensity within an arbitrarily determined speaking range. Two ways of selecting these frequencies are proposed here.

The first one is independent of the absolute frequency scale and uses information on the individual mean speaking fundamental frequency (mff). In the other procedure standardized male and female mff's are used. The mff of male subjects was set at 123 Hz, while a mff of 220 Hz was chosen for female subjects (see also Awan, (27), using mff's of 123 Hz and 206.6 for male and female subjects, respectively). In both procedures the other three frequencies at which the dynamic range is investigated are: three semitones below mff, half an octave and an octave above mff. We assumed that with these frequencies the speaking voice range is largely covered.

With a microphone at a distance of 30 cm from the mouth, measured mean intensities of normal speech will generally fall between 60 and 80 dB (27). This intensity range is therefore most important for a normal production and

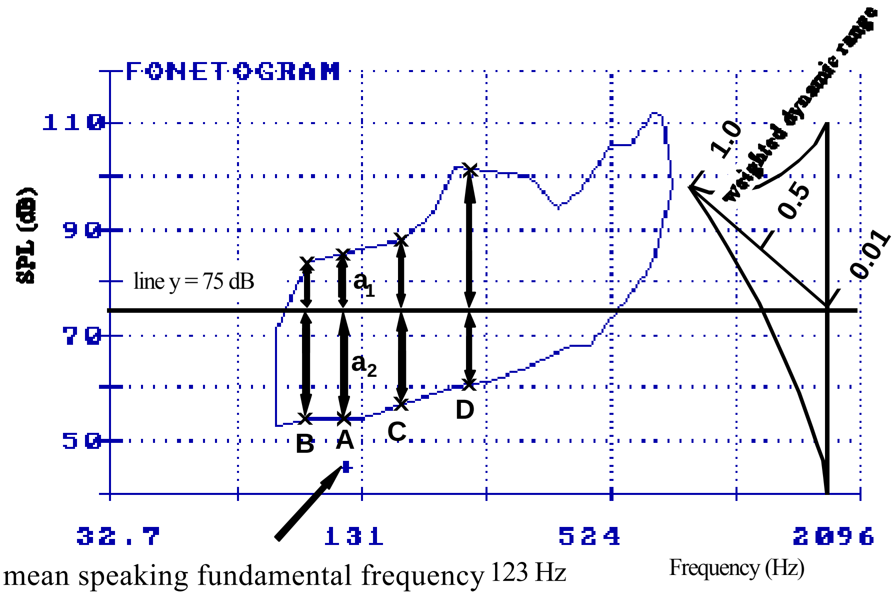


Figure 6. To assess "Speaking Range" dynamics at four frequencies (A, B, C and D) the distance from the reference line of 75 dB to the loud phonation contour (a_1) and the soft phonation contour (a_2) is determined. The four frequencies are: the mean speaking fundamental frequency (mff) (A), mff minus 3 semitones (B), mff plus half an octave (C) and mff plus an octave (D). Because of the relative importance in normal speech of intensities around 75 dB the distances are processed with a weighting factor. In the figure the weighting factor is indicated in the third dimension. With the distances from the reference line and the imposed weighting factor the center of the dynamic range can be determined.

communication of running speech. A restricted intensity range may affect intonation and stress patterns and thus reduce the quality of spoken language. The intensities above 80 dB and below 60 dB normally will not be used during speech; however, the ability to raise one's voice is necessary for adapting to these special occasions which demand high intensities.

Because all intensities are not equally important during normal speech production a weighting factor was introduced in calculating the intensity range at the given frequencies. The weighting factor uses the natural logarithm of the measured values. It enhances the importance of intensities used in normal speech, in contrast to extremes in vocal loudness that are only used occasionally in shouting (loud voice) or quiet conversation (soft voice). The line representing an intensity of 75 dB is arbitrarily selected as the reference intensity for a normal intensity modulation. On both sides of the 75 dB line the importance of the intensity decreases approximating the decay of a natural logarithm. Figure 6 gives the selected frequencies together with a graphical illustration of the weighting factor.

Weighted dynamic range. At the four frequencies the distances (in dB) of the measured minimum and maximum intensity from 75 dB are calculated. When a minimum value of, for instance, 55 dB is measured, the distance from the reference line is 20 dB. This relative value is processed with the weighting factor; that is, the natural logarithm is taken of the value, resulting in a new weighted value of 3.0. The same procedure applied to a maximum intensity of, for instance, 85 dB results in a weighted value of 2.3. The weighted dynamic range (55 - 85 dB) thus gives a value of 3.0 + 2.3 = 5.3. As a result of this weighting procedure, the maximal value will be obtained when the 75 dB line passes through the midpoint of the dynamic range.

Central position. The central position of the range is obtained by adding the weighted minimum and maximum distances (from 75 dB) and dividing the sum by two. The minimum and maximum intensities of 55 and 85 dB in the preceding example yield a central position of (-3.0 + 2.3) / 2 = -0.35. The negative sign indicates a central position of the dynamic range below the reference intensity of 75 dB.

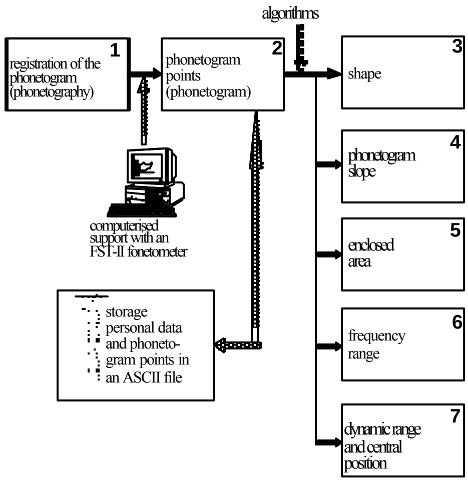


Figure 7. The process diagram of phonetography and analysis of registered phonetograms. With the aid of a computerized phonetometer (1) the phonetogram points are acquired. Together with personal data these points are stored on floppy disk (2) for archival purposes and for further analysis. With algorithms incorporated in an application program the features shape (3, 4), area (5, 6), and weighted dynamic range and central position at selected frequencies (7) are analyzed.

Application of the analytical procedure

The application of this structured analysis of phonetograms with algorithms, yielding parameter values, was implemented in a computer program. The application program was written in a fourth generation signal analysis software package ASYST (MacMillan Software Company), which operates in DOS environment.

Data files proceeding from the computerized registration of phonetogram points consist of a header containing personal data, followed by a clockwise listing of phonetogram points. This file can be used to generate a phonetogram or to serve as the input for the application program. Figure 7 gives a process diagram that summarizes the analytic procedures performed.

Processing multiple phonetograms can be done easily by using standardized

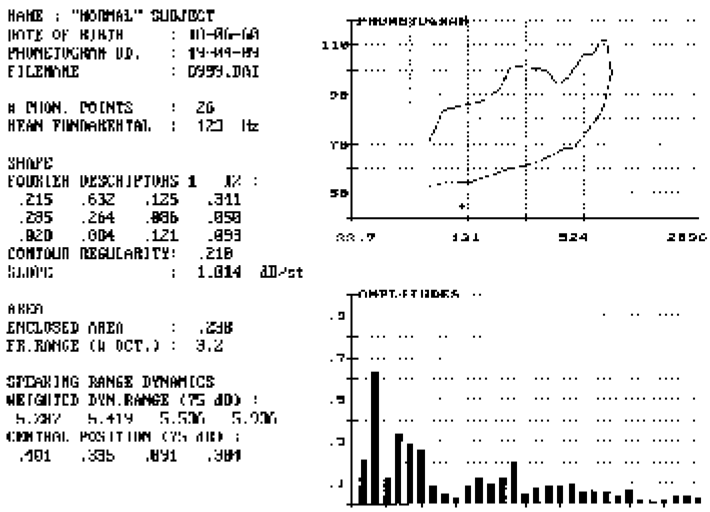


Figure 8. The result of the analytic computation is displayed on a monitor. In the upper-right corner the phonetogram is plotted. Underneath the first 30 Fourier Descriptors are given. At the left side the personal data and analyzed parameters are printed.

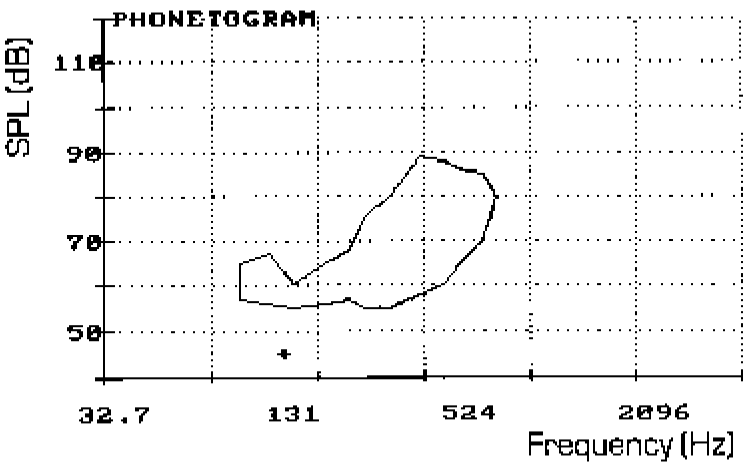


Figure 9. The phonetogram of a male subject with a mutational voice disorder.

filenames with a numeric part corresponding to a specific subject or patient. This makes possible a highly automated evaluation of a large number of phonetograms when the input for analysis is specified by a starting and ending file number. Figure 8 shows the output displayed on the computer screen. The resulting parameters are displayed on the monitor and stored in an output file. Combining multiple individual output files originating from analyzed phonetograms of subjects without voice complaints, reference parameter values for persons belonging to this group can be established.

Illustration of the application program

To demonstrate the functioning of the program and to illustrate its capability in determining parameters, phonetograms of two male subjects, one without and one with a mutational voice disorder, were processed. The resulting parameters are compared with the mean values and standard deviation of a large group (n=46) of male subjects without voice complaints or voice training, hereafter referred to as the normal reference group (24).

Figure 9 gives the phonetogram of the male subject with a mutational voice disorder, while the phonetogram of the "normal" male subject is used for illustrating the proposed method (see Figures 1 to 8). As far as function is concerned the most salient abnormal aspect of the phonetogram displayed in

Parameter	Normal	Pathologic	Mean	SD
Shape				
FD_1	.22	.33	.14	.089
FD_2	.63	.84	.57	.089
FD_3	.13	.36	.23	.119
FD_4	.34	.32	.25	.107
FD_5	.29	.07	.20	.129
FD_6	.26	.17	.24	.101
FD_7	.09	.06	.16	.081
FD_8	.05	.10	.18	.094
FD_9	.03	.10	.14	.075
FD_{10}	.08	.18	.16	.098
FD_{11}	.12	.04	.13	.071
FD_{12}	.09	.09	.12	.075
Contour Regularity	.21	.17	.21	.034
Slope (dB/st)	1.01	.64	.91	.208
Area				
Enclosed area	.24	.11	.25	.052
Frequency range (# octaves)	3.2	2.8	3.2	.32
"Speaking Range" Dynamics				
$WDR_{mff-3\ st}$	5.3	0.8	3.4	1.84
WDR_{mff}	5.4	0.5	5.2	1.10
$WDR_{mff+1/2\ oct}$	5.5	0.7	6.0	.41
$WDR_{mff+1\ oct}$	5.9	4.2	6.0	.40
$Position_{mff-3\ st}$	-0.4	-2.5	-1.6	.91
$Position_{mff}$	-0.3	-2.8	-0.7	.54
$Position_{mff+1/2\ oct}$	-0.1	-2.6	-0.2	.21
$Position_{mff+1\ oct}$	0.3	-0.9	0.1	.19

Table 1. Parameter values for a "normal" and a pathological (mutational voice disorder) phonetogram with reference values (mean and Standard Deviation [SD]) of a group of 46 male subjects without vocal complaints or vocal training.

Figure 9 is the restricted dynamic range at lower frequencies. To indicate specific deviations from "normal" male phonetograms, both phonetograms are analyzed with the mff standardized at 123 Hz (B2 in American notation).

Table 1 gives the analyzed parameter values of the "normal" and pathologic phonetograms, as well as the mean values and standard deviations (SD) of the normal reference group. The "normal" phonetogram yields parameter values all within a range of 2 SDs from the mean value, which is commonly accepted as defining a normal range. The subject with a mutational voice disorder, however, produced a phonetogram that yields deviant parameter values. Regarding the shape, the first and second FD show values above the range of 2 SDs. The enclosed area is small compared to the reference norm. Summarizing the parameters of the "speaking range" dynamics, the weighted ranges are, except for the lowest sampled frequency, all significantly small and the central positions of these ranges are well below the reference intensity, which means that at all sampled frequencies phonations are only possible with soft intensities. In short, the phonetogram of the subject with vocal problems is abnormal with respect to shape, enclosed area and "speaking range" dynamics, and has parameter values that might conceivably be representative for a mutational voice disorder.

DISCUSSION

The power and robustness of the proposed parameters largely depend on a standardized registration of phonetogram points. Directions and instructions were formulated by Schutte & Seidner (6). However, because the shape parameters are dependent on the number of points in a phonetogram, we strongly advise a consistent choice of points at which the frequency range is sampled. Following Schutte & Seidner, four frequencies per octave are recommended. When a recommended step on either end of the range is beyond the phonatory possibilities of the tested subject, an increment or reduction by a tone or semitone will provide the extremes.

Fourier Desciptors can be used to describe quantitatively a shape. Applying the analytic method proposed by Zahn and Roskies (26) to phonetograms results in an order of Fourier Descriptors with varying amplitudes (see Figure 2c). By processing a large number of phonetograms and averaging amplitudes a "mean" phonetogram can be obtained. Further research is needed to establish specific relationships between one or more Fourier Descriptors.

The shape is influenced by the dB(A) weighting network used to register the sound level (18). The increasing attenuation of frequencies below 500 Hz reduces the SPL levels of the loud and especially the soft contours at the low frequencies. Where phonetogram analysis is used as a method for comparing phonetograms and for observing and detecting changes in phonetograms under the influence of therapy or training, its power as a clinical instrument is not compromised by a standardized use of a dB(A) weighting network. However, this weighting network does limit the scientific use of phonetograms in research on voice function.

Registering phonetograms we accept reproducible phonations at a given frequency with a minimum phonation time of one second. This duration insures a stabilized sound intensity production and correct measurement. A smaller minimum phonation time, as compared with the three seconds recommended by Coleman, Mabis, & Hinson (11) and Schultz-Coulon & Asche (3), reveals the physiological extremes of the voice and makes the procedure more practicable in clinical practice with patients suffering from laryngeal diseases associated with short phonation times.

The description of phonetograms with explicit parameters offers the possibility of determining deviations, from normal values, for pathological entities. If these deviations show specific disease-related characteristics, each type of disease with effects upon phonation and phonatory potentialities could be represented by a set of parameter statistics. Knowledge about a special combination of parameter statistics can be used to build an expert system (28). In an expert system specific knowledge on observed behavioral patterns is formalized in a computer program, giving the users a stronger basis for making decisions. A suggested expert system based on knowledge of constellations of parameter values of phonetograms, specific for groups of subjects or patients, could be a very useful tool for speech pathologists, therapists and ENT-physicians to help in diagnosing diseases or in supporting a diagnosis.

Before this can be realized, a large number of phonetograms per pathologic entity have to be analyzed in order to build the knowledge base. A computerized phonetometer with analysis techniques can make a major contribution toward an optimal clinical employment of phonetograms.

References

1. Calvet P, Malhiac G. Courbes vocales et mue de la voix. J Franç d'Otorhinolaryngol Chir Maxillofac 1952;1:115-24.
2. Klingholz F, Martin F. Die quantitative Auswertung der Stimmfeldmessung. Sprache-Stimme-Gehör 1983;7:106-10.
3. Schultz-Coulon H, Asche S. Das "Normstimmfeld" - ein Vorschlag. Sprache-Stimme-Gehör 1988;12:5-8.
4. Waar C, Damsté P. Het fonetogram. Logopedie en Foniatrie 1968;40: 198-201.
5. Schutte H. Over het fonetogram. Logopedie en Foniatrie 1975;47:82-92.
6. Schutte H, Seidner W. Recommendation by the Union of European Phoniatricians (UEP): Standardizing Voice Area Measurement/Phonetography. Folia Phoniatr 1983;35:286-8.
7. Pabon J, Plomp R. Automatic phonetogram recording supplemented with acoustical voice-quality parameters. J Speech Hear Res 1988;31:710-22.
8. Bloothooft G. Nieuwe ontwikkelingen in de fonetografie. Logopedie en Foniatrie 1982;54:79-90.
9. Jaroma M, Sonninen A, Hurme P, Toivonen R. Computer voice field observations of postmenopausal dysphonia. In: Frank F, ed. Proceedings of the XIIIth International Congress of the Union of European Phoniatricians Vienna. Vienna, Austria: UEP, 1986:34-5.
10. Wolf S, Stanley D, Sette W. Quantitative studies on the singing voice. J Acoust Soc Am 1935;6:255-66.
11. Coleman R, Mabis J, Hinson J. Fundamental frequency-sound pressure level profiles of adult male and female voices. J Speech Hear Dis 1977;20:197-204.
12. Damsté P. The phonetogram. Practica Otorhinolaryngol 1970;32:185-7.
13. Dejonckere P. Le Phonetogramme, son intérêt clinique. Les Cahiers d'Oto-Rhino-Laryngologie 1977;12:865-72.
14. Hollien H, Dew D, Philips P. Phonational frequency ranges of adults. J Speech Hear Res 1971;14:755-60.

15. Hirano M. Clinical examination of voice. Bd.5. In: Arnold G, Winckel F, Wyke B, eds. Disorders of Human Communication. Wien, New York: Springer, 1981.
16. Schutte H. Untersuchungen von Stimmqualitäten durch Phonetographie. HNO Praxis 1980;5:132-9.
17. Schutte H. Belastbaarheid van de stem en het fonetogram. Logopedie en Foniatrie 1986;58:140-3.
18. Gramming P, Sundberg J. Spectrum factors relevant to phonetogram measurement. J Acoust Soc Am 1988;83:2352-60.
19. Vilkman E, Sonninen A, Hurme P. Observations on voice production by means of computer voice fields. In: Kirikae I, ed, Proceedings of the XXth International Congress of Logopedists and Phoniatricians IALP Tokyo. Tokyo, Japan: Organizing Committee of the XXth Congr IALP, 1986:370-1.
20. Titze I. Acoustic interpretation of the voice range profile (phonetogram). J Speech Hear Res 1992;35:21-34.
21. Åkerlund L. (1993). Averages of sound pressure levels and mean fundamental frequencies of speech in relation to phonetograms: comparison of nonorganic dysphonia patients before and after therapy. Acta Otolaryngol 1993;113:102-8.
22. Åkerlund L, Gramming P, Sundberg J. Phonetogram and averages of Sound Pressure Levels and fundamental frequencies of speech: Comparison between female singers and nonsingers. J Voice 1992;6:55-63.
23. Awan S. Phonetographic profiles and F0-SPL characteristics of untrained versus trained vocal groups. J Voice 1991;5:41-50.
24. Sulter A, Schutte H, Miller, D. Differences in phonetogram features between male and female subjects with and without vocal training. J Voice 1995;9:363-77.
25. Heinemann M, Gabriel H. Möglichkeiten und Grenzen der Stimmfeldmessung - Vorstellung Heiserkeitsfeldes als Ergänzung der Methode. Sprache-Stimme-Gehör 1982;6:37-42.
26. Zahn C, Roskies R. Fourier Descriptors for plane closed curves. IEEE Trans Comp 1972;21:269-81.
27. Awan S. Superimposition of speaking voice characteristics and phonetograms in untrained and trained vocal groups. J Voice 1993;7:30-7.
28. Keller R. Expert system technology. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1987.

8

Differences in phonetogram features between male and female subjects with and without vocal training

Arend M. Sulter, Harm K. Schutte [*] and Donald G. Miller [*]

Department of Otorhinolaryngology, University Hospital Groningen and [*]Voice Research Lab, Department of Medical Physiology, Faculty of Medicine, University of Groningen, Groningen, the Netherlands

Previously published in the *Journal of Voice* 1995,9(4),363-377.

INTRODUCTION

Vocal capabilities vary among persons, depending on individual laryngeal anatomy, as well as physical and physiological conditions. This variation presents a continuum from poor to excellent vocal function. In order to place a given voice along this continuum some quantitative measures of vocal function have to be used. Two conveniently measurable quantities, substantially characterizing the acoustic output of the voice, are Sound Pressure Level (SPL) and fundamental frequency (F_0). Plotting minimum and maximum SPL along a person's frequency range, the voice range profile (phonetogram) can be determined (1-3). The resulting phonetogram exhibits the vocal capabilities of a person and provides information on the possibilities of the voice source, as well as the selective amplification of the vocal tract (4-7).

To evaluate vocal function it is necessary to have normative data on human vocal capabilities. Such data, as displayed in a phonetogram, appear to be scarce (8,9). These data are necessary for qualifying vocal function. A key to the establishment of normative databases is the use of clearly defined criteria for the inclusion of subjects. Such clearly defined criteria will also allow the creation of subsets of databases.

One obvious criterion for a subset is gender, since a number of studies have established the basic difference in laryngeal anatomy and physiology between men and women (10-12). Another criterion is vocal training. Studies have shown, for example, the different way singers adjust the vertical larynx position compared to untrained subjects (13), and the specific respiratory movements singers apply during the act of singing (14). Trained subjects, who are accustomed to exploiting more fully their dynamic and pitch ranges, might be expected to produce phonetograms with larger capacities. If this is indeed the case the results of the trained group might give a perspective on the interpretation of the poor to excellent continuum, offering information about the excellent side of this continuum.

In this study normative data of phonetograms are established for untrained, as well as trained, male and female subjects. Gender-related differences will be analyzed along with differences related to vocal training. Finally, phonetogram variables that best distinguish between the trained and the untrained will be determined.

METHODS

Subjects

A total of 224 Dutch untrained and trained subjects of both genders, categorized accordingly into 4 groups, were investigated. The untrained subjects were recruited from groups of students and volunteers without vocal complaints or history of vocal pathology. The group consisted of 92 females and 47 males. The mean age for the female subjects in this subgroup was 20.3 years, ranging from 17 to 44, while the mean age for the male subjects was 25.0 years, ranging from 17 to 35.

42 female and 43 male subjects with a minimum of two years of vocal training served as another group. The vocal training could either consist of singing in a choir that organized conducted rehearsals with a minimum frequency of once a week, or receiving individual singing lessons with a similar minimum frequency. The mean age of the female trained group was 35.1 years, ranging from 18 to 59, and the mean age of the male subjects was 47.5 years, ranging from 21 to 75.

Equipment

Phonetograms were registered in a sound treated room with 'living room acoustics' (15), using an FST-II phonetometer. This equipment consists of an Atari ST computer, additional hardware, application software and a Monacor® omnidirectional dynamic microphone. During operation the computer screen shows the template of a phonetogram. Sound Pressure Level is measured with a "fast" sound level meter, containing a dB(A) weighting network (high pass to reduce measuring low frequency noise). A mouse connected to the computer moves an arrow on the screen. Pressing the left mouse button, the location of this arrow along the frequency scale determines the pitch of a vowel-like tone, which is used as a prompt tone for the subject. Pressing the right button, the intensity level produced is captured and a cross appears on the screen at the selected frequency and measured SPL. After mapping the whole frequency range for minimum and maximum sound intensities the phonetogram points are stored on diskette. A phonetogram is created by connecting the points with line segments.

All phonetograms were registered by two investigators, both familiar with the equipment and phonetographic procedure, and having had adequate musical training and experience. In a few test sessions they were trained to use a standard set of instructions for the subjects. Subjects from the four groups were randomly assigned to the two investigators.

Phonetographic procedure

Phonetography recommendations by the UEP were followed (15). The direction and distance (30 cm) of the subject's mouth to the microphone was carefully controlled during the procedure. The subjects were tested using the vowel /a/. Phonetogram points were collected, starting the acquisition at the mean speaking fundamental frequency, followed by the low frequencies and ending with high frequencies. The mean speaking fundamental frequency was determined by asking the subjects to count from one to ten. Then the investigator's imitation of that frequency was measured with the phonetometer. A reproducible phonation at a given frequency with a minimum phonation time of one second, to insure a stabilized sound intensity production and correct measurement, was required for accepting a phonetogram point. During actual registration the subject, if necessary, was at first guided in matching the target frequency, and after this the minimum and maximum intensity were registered. The subjects were instructed to produce phonations at the physiologic boundaries, without, of course, injuring the voice during maximum intensity. The frequency range was sampled at four frequencies per octave, basically at the tones c-e-g-a, e.g., C3-E3-G3-A3 (in this octave 131, 165, 196 and 220 Hz). At the upper and lower ends of the range shorter frequency intervals were chosen. The phonetograms were acquired in ten to twenty minutes. Not surprisingly, testing untrained subjects generally required more time than registering vocal capabilities in trained subjects.

Phonetogram analysis

For reasons of comparison and establishing differences between phonatory capacities a standardized approach was used. Phonetograms were analyzed

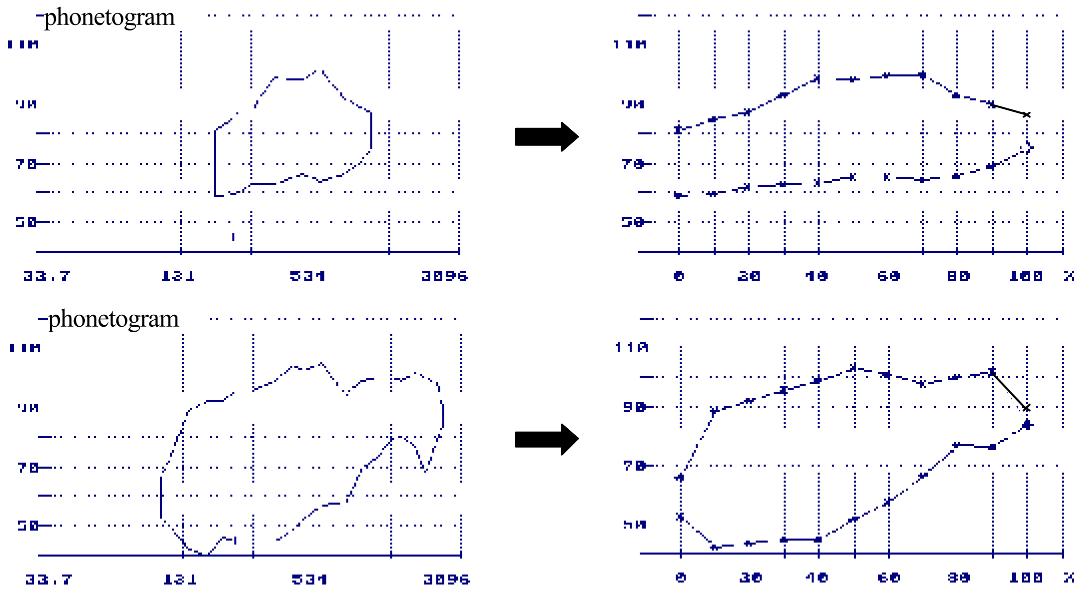


Figure 1. Rescaling phonetograms and the effect on shape. On top a small phonetogram is stretched and below loss of detail of the phonetogram can be observed.

using two different methods.

Rescaling method. The rescaling method has been described and used in many published studies (6,7,16-18). It determines dynamic ranges at fixed relative distances along a subject's (individually differing) frequency range. Retrieving phonetogram points from individual datafiles, each frequency range was rescaled to 100% with a specially developed computer program. At 10% intervals the minimum and maximum intensity was calculated by interpolation, yielding eleven values for both minimum and maximum sound pressure levels. By averaging the SPL values at each interval for a number of subjects, normative data on the dynamic ranges were established. However, though this method supplies data on dynamic ranges, it lacks a consistent relation with the absolute frequency scale and introduces a considerable distortion of the shape of phonetograms.

Figure 1 depicts this process of distortion by contrasting two phonetograms, one with a small, and one with a with a large frequency range. After rescaling to 100% on a standard template the upper phonetogram is stretched, thereby dramatically changing the shape. The lower phonetogram shows loss of detail of the shape. Because of the loss of the absolute frequency-related information, the originally measured datapoints cannot be read from such a rescaled phonetogram. However, the advantage of this procedure is that phonetographic data can be treated and compared numerically.

Conjoint frequency and intensity analysis of the phonetogram The authors have presented a new approach for the analysis of phonetograms, concentrating on the phonetogram features shape, area and weighted dynamic range and central position. The main difference between this new approach and the rescaling method is that it derives variables without distortion of the shape. What follows is a brief description of this new approach. For a detailed description the reader is referred to Sulter et al. (9).

The shape of a phonetogram is described with so-called Fourier Descriptors, thus enabling quantification (19). In short, this complex procedure translates directional angle of connecting lines between phonetogram points into information that can be processed with a Fourier transformation. The amplitudes of Fourier Descriptors with a low number represent the general shape whereas the descriptors with higher number are related to small changes in shape. Contour regularity also describes shape and is given as the quotient of enclosed area and squared perimeter, producing a dimensionless value. The slope of a phonetogram is determined by drawing a line with minimal distance to all points through the phonetogram and calculating the angle between this line and x-axis.

The enclosed area gives the quotient of the voice field area and the rectangle with coordinates 40 and 110 dB, and 32.7 and 2096 Hz. This relative area makes the value independent of plotting factors (20) and thereby allows comparisons with results from other studies. The frequency range is derived by subtracting the lowest phonated fundamental frequency (F_0) from the highest frequency.

A weighted dynamic range (WDR) is determined in the modal register at four frequencies[1] which are related to the mean speaking fundamental frequency. Because sound intensities around 75 dB are relatively more important in normal speech than very soft and loud intensities, a logarithmic weighting factor was used, giving greater weight to the intensities in accordance to how close they are to 75 dB. Besides the WDR the central position (CP) of this range is calculated.

Both these methods for analyzing phonetograms were implemented in a computer program written in the ASYST language (ASYST, Macmillan software company) and running in DOS environment. Resulting output datafiles were processed with Quattro spreadsheet software (Borland

---

[1]The four sampled frequencies in the modal register are: mean speaking fundamental frequency (mff) minus three semitones; mff; mff plus half an octave; and mff plus an octave. In this study the mff was standardized at 220 Hz for female subjects and 123 Hz for male subjects.

International, Inc.), and SPSS statistical software (SPSS, Inc.).

Statistical analysis

To analyze differences between phonetograms, many comparisons and related statistical tests were applied to those variables that resulted from employing both the rescaling and conjoint methods. Because many tests are performed the probability level α was set at 0.001 in order to minimize the chance (type I error) of erroneously rejecting the null hypothesis, which posits no difference among variables under investigation. This conservative probability level was established with respect to the Bonferroni inequality. The large number of subjects assured adequate power (probability of rejecting the null hypothesis when it is false) of the tests.

Two-way analysis of variance, with gender and vocal training as factors, was used to establish compactly differences between groups. Only when a significant interaction between gender and vocal training occurred for a variable was a separate T-test performed on each factor level.

RESULTS

Results are presented below according the two methods of phonetogram analysis. Comparisons were made between male and female, as well as between untrained and trained subjects.

Rescaling method

After averaging a large number of phonetograms with the rescaling method descriptive statistics can be derived for each frequency level and vocal intensity. These averaged frequency levels will first be discussed for all analyzed groups, and then the averaged minimum and maximum intensity levels will be presented separately for female and male subjects. After averaged phonetograms are presented, inferential statistics are used to establish differences in phonetograms between groups.

Frequency levels. Table 1 gives averaged frequencies with standard deviations at the 10% frequency levels for untrained and trained groups.

An average frequency range from 157.3 Hz to 1223.7 Hz was measured in untrained female subjects. The trained counterpart exhibited frequency values ranging from 128.4 Hz to 1320.3 Hz. Thus the trained female subjects exceeded their untrained counterparts at both ends of the frequency range. Trained male subjects also exceeded their untrained counterparts at the 0% frequency level

		Female		Male	
Frequency level		Untrained (n=92)	Trained (n=42)	Untrained (n=47)	Trained (n=43)
0%	Mean (Hz)	157.3	128.4	86.1	74.0
	SD (Hz)	21.42	17.62	14.01	13.49
10%	Mean (Hz)	192.6	161.7	107.1	92.0
	SD (Hz)	23.44	20.51	16.08	15.86
20%	Mean (Hz)	235.9	203.5	133.3	114.5
	SD (Hz)	26.24	24.74	18.90	19.45
30%	Mean (Hz)	289.3	256.6	166.0	142.7
	SD (Hz)	30.59	31.29	23.01	24.91
40%	Mean (Hz)	354.7	323.5	206.9	178.0
	SD (Hz)	37.85	41.36	29.17	33.14
50%	Mean (Hz)	435.2	408.3	258.1	222.5
	SD (Hz)	49.48	56.73	38.37	45.16
60%	Mean (Hz)	534.5	515.5	322.0	278.2
	SD (Hz)	67.18	78.90	51.74	62.30
70%	Mean (Hz)	656.9	651.5	401.9	348.3
	SD (Hz)	93.06	110.85	70.99	86.30
80%	Mean (Hz)	807.8	824.0	502.4	436.6
	SD (Hz)	129.59	155.84	98.40	119.86
90%	Mean (Hz)	993.8	1042.9	628.0	547.9
	SD (Hz)	180.32	218.27	136.24	165.73
100%	Mean (Hz)	1223.7	1320.3	785.4	688.2
	SD (Hz)	249.44	303.93	188.38	228.14

Table 1. Mean frequencies and standard deviations (SD) for consecutive frequency levels of untrained and trained female and male groups.

with 74.0 Hz versus 86.1 Hz, respectively. However, at the 100% frequency level the untrained male subjects showed a higher average frequency of 785.4, as compared with a frequency of 688.2 Hz in the trained males.

The standard deviation shows an absolute increase in hertz as the frequency level rises. In semitones this amounts to about 2 semitones at the 0% frequency levels and 3 semitones at the 100% frequency levels. Male subjects (especially

Frequency level		Maximum Untrained (n=92)	Maximum Trained (n=42)	Minimum Untrained (n=92)	Minimum Trained (n=42)
0%	mean (dB)	63.1	55.9	53.9	49.6
	SD (dB)	10.49	10.11	6.86	8.61
10%	mean (dB)	82.4	77.8	51.4	44.2
	SD (dB)	6.72	7.65	6.52	4.35
20%	mean (dB)	90.0	87.9	53.4	46.0
	SD (dB)	5.23	4.95	6.60	4.51
30%	mean (dB)	94.5	93.3	56.1	49.2
	SD (dB)	5.28	4.54	6.47	4.75
40%	mean (dB)	97.4	96.1	58.4	52.8
	SD (dB)	5.61	4.47	6.30	4.64
50%	mean (dB)	97.3	97.6	61.0	56.2
	SD (dB)	6.38	6.11	6.71	5.49
60%	mean (dB)	97.5	99.5	65.2	60.8
	SD (dB)	5.73	5.26	7.40	5.27
70%	mean (dB)	99.6	101.4	70.6	66.2
	SD (dB)	5.54	5.99	7.74	6.08
80%	mean (dB)	102.0	102.9	76.8	71.7
	SD (dB)	6.02	7.06	8.60	7.24
90%	mean (dB)	103.8	104.1	83.6	80.2
	SD (dB)	6.30	7.75	10.24	9.41
100 %	mean (dB)	102.7	99.9	97.7	94.5
	SD (dB)	8.73	9.76	11.47	11.95

Table 2. Maximum and minimum intensities in untrained and trained female subjects. Mean intensity levels and standard deviations (SD) are given at 10% frequency levels .

trained males) show the largest distribution in frequency values. Thus the male trained group shows the largest differences among phonetograms with respect to both frequency range and frequency position.

Frequency level		Maximum Untrained (n=47)	Maximum Trained (n=43)	Minimum Untrained (n=47)	Minimum Trained (n=43)
0%	mean (dB)	59.2	59.8	52.0	52.4
	SD (dB)	10.15	6.93	7.69	6.92
10%	mean (dB)	77.2	74.9	46.9	46.2
	SD (dB)	6.71	7.81	6.02	5.42
20%	mean (dB)	85.7	83.9	46.6	45.7
	SD (dB)	4.76	6.38	5.74	4.43
30%	mean (dB)	90.7	90.6	48.7	47.5
	SD (dB)	4.66	5.36	5.23	5.12
40%	mean (dB)	94.8	95.1	52.0	50.1
	SD (dB)	5.03	4.37	5.68	4.99
50%	mean (dB)	97.9	98.6	56.9	54.3
	SD (dB)	5.35	4.44	6.21	4.77
60%	mean (dB)	100.3	100.6	61.5	58.3
	SD (dB)	5.57	4.84	7.23	5.67
70%	mean (dB)	98.0	101.2	65.3	62.9
	SD (dB)	5.92	4.91	8.05	6.27
80%	mean (dB)	97.3	100.7	71.2	68.6
	SD (dB)	7.31	6.47	8.72	8.06
90%	mean (dB)	99.7	100.7	77.9	75.3
	SD (dB)	8.07	8.54	9.23	11.70
100%	mean (dB)	96.4	99.2	92.0	90.6
	SD (dB)	11.19	10.8	12.22	14.46

Table 3. Maximum and minimum intensities in untrained and trained male subjects. Mean intensity levels and standard deviations (SD) are given at 10% frequency levels.

Sound intensity levels. Female subjects. Table 2 gives the averaged maximum and minimum phonation intensities with standard deviations at the 10% frequency levels, together with the standard deviation for the female groups of untrained and trained subjects. With the averaged frequencies of Table 1, Figure 2 was constructed, visualizing the averaged phonetograms for untrained and trained female subjects.

The standard deviations for frequencies and intensities are represented as

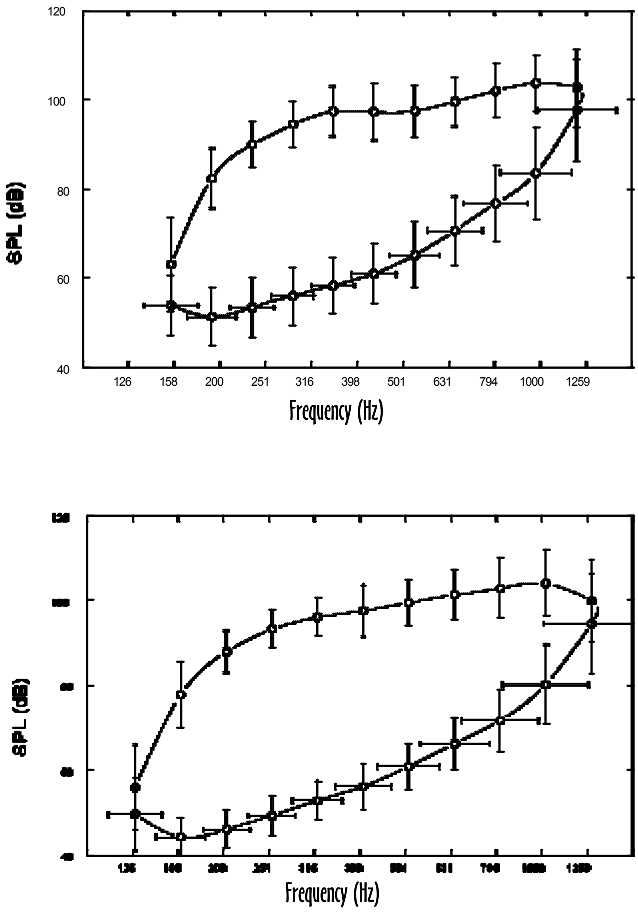


Figure 2. Averaging frequency and intensity data mean phonetograms are plotted for a group of 92 untrained (above) and 42 trained female subjects (below). The whiskers indicate 1 standard deviation.

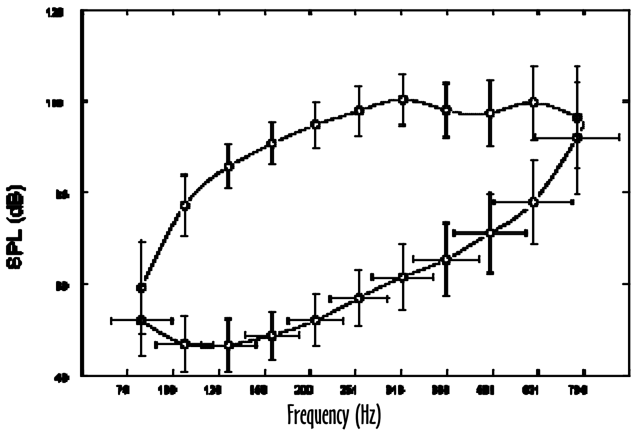
whiskers in both plots. The whiskers show, especially at higher frequencies, a large overlap, due to the large standard deviations for these frequency values. Apart from the values near the extremes, the standard deviations of intensities are generally less than 7 dB.

With increasing frequency a rising intensity level is present for the soft phonation contour in both the untrained and trained female phonetogram. The averaged phonetogram of the trained female group shows, generally, better possibilities for producing soft intensities, compared to the untrained counterpart.

A distinct difference in the loud phonation contour between the averaged phonetograms is a relative minimum in an otherwise continuously rising loud phonation contour, present at the 50-60% frequency levels (about 400-500 Hz) of the averaged untrained phonetogram, whereas this is virtually absent in the trained phonetogram. It marks the transition from modal to falsetto register, which therefore seems to occur in a more smoothly way in the trained female group.

Male subjects. Table 3 gives the averaged maximum and minimum phonation intensities at the frequency levels, together with the standard deviation, for the male groups of untrained and trained subjects. With averaged frequency values of Table 1, plots of averaged phonetograms for untrained and trained male subjects were made (see Figure 3).

The averaged phonetograms of untrained and trained males show a high degree of similarity; however, there are important differences. Trained males have greater voice capacities in the lower frequency region, while the



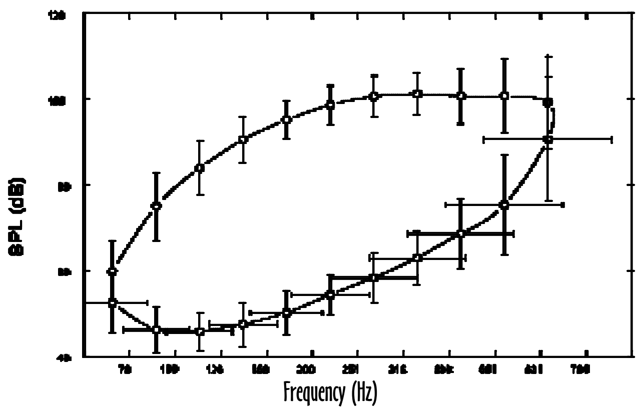


Figure 3. Averaging frequency and intensity data mean phonetograms are plotted for a group of 47 untrained (above) and 43 trained male subjects (below). The whiskers indicate 1 standard deviation.

phonations in falsetto (those above about ca. 70% frequency level) are limited, particularly with respect to frequency range. Another important point of difference is in the transition from modal to falsetto voice: while the untrained male phonetogram shows an evident depression at the 70-80% frequency level (about 400-500 Hz), the smooth contour of loud intensities is maintained in the trained counterpart.

Differences between phonetograms Table 4 gives the results of the ANOVA procedure applied to the consecutive frequency levels. No significant interaction is present between the factors gender and training. Not surprisingly, the effect of the factor gender on frequency is significant at all frequency levels. The factor vocal training shows a significant effect only at those

	Gender		Vocal training		Gender x training	
	F	p	F	p	F	p
FR0	672.16	<0.001 *	73.60	<0.001 *	10.97	0.001
FR10	803.68	<0.001 *	72.63	<0.001 *	7.79	0.006
FR20	905.16	<0.001 *	65.73	<0.001 *	4.31	0.039
FR30	935.42	<0.001 *	52.45	<0.001 *	1.40	0.238
FR40	869.07	<0.001 *	35.98	<0.001 *	0.05	0.825
FR50	739.99	<0.001 *	20.99	<0.001 *	0.41	0.523
FR60	599.35	<0.001 *	10.59	0.001	1.78	0.183
FR70	472.79	<0.001 *	4.30	0.039	3.53	0.062
FR80	369.69	<0.001 *	1.23	0.270	5.23	0.023
FR90	291.20	<0.001 *	0.09	0.763	6.70	0.010

Table 4. Analysis of variance summary table for frequency levels.
Note. df = 1, 222 for gender, vocal training, and gender x training. The figure behind FR indicates the frequency level, e.g., FR80 means the frequency level 80%. *p < 0.001

frequency levels from 0% to 60%. In this range trained subjects at similar frequency levels phonate at significantly lower frequencies compared to untrained subjects.

Differences in intensities between groups were also compared. Table 5 gives the results of the ANOVAs that were performed on the minimum intensity values. A significant interaction was found between the factors gender and training at the frequency levels 10 and 20%. Separate T-tests were performed at these levels. A significant effect of the factor gender was found at the frequency levels 30, 40 and 50%, as well as 70, 80 and 90%. At all these levels male subjects are able to phonate more softly than female subjects.

The factor vocal training showed a significant effect at the frequency levels 30, 40, 50 and 60%. Trained subjects are able to phonate more softly at these frequency levels.

T-tests performed within the untrained subgroup showed a significant difference between male and female subjects at the 10 and 20% frequency level ($t = 3.97$, $df = 137$, $p < 0.001$, and $t = 5.99$, $df = 137$, $p < 0.001$, respectively). Male subjects were able to phonate with softer intensities at these frequency levels. Within the trained subgroup no significant differences were found at these frequency levels ($t = -1.82$, $df = 83$, $p = 0.072$, and $t = 0.25$, $df = 83$, $p = 0.800$, respectively). Within the female subgroup significant differences were found between untrained and trained subjects at the 10 and 20% frequency level ($t = 6.50$, $df = 132$, $p < 0.001$, and $t = 6.64$, $df = 132$, $p < 0.001$, respectively), with trained subjects able to phonate at softer intensities. T-tests performed within the male subgroup showed no significant differences ($t = 0.59$, $df = 88$, $p = 0.560$, and $t = 0.83$, $df = 88$, $p = 0.407$, respectively).

	Gender		Vocal training		Gender x training	
	F	p	F	p	F	p
DYNL0	0.00	0.988	4.78	0.030	5.32	0.022
DYNL10	5.50	0.020	28.26	<0.001 *	15.41	<0.001 *
DYNL20	27.64	<0.001 *	33.29	<0.001 *	16.69	<0.001 *
DYNL30	41.66	<0.001 *	31.02	<0.001 *	12.17	0.001
DYNL40	38.51	<0.001 *	25.22	<0.001 *	5.20	0.023
DYNL50	14.48	<0.001 *	20.78	<0.001 *	1.74	0.189
DYNL60	11.92	0.001	16.93	<0.001 *	0.34	0.558
DYNL70	20.07	<0.001 *	12.39	0.001	0.99	0.322
DYNL80	16.13	<0.001 *	11.89	0.001	1.17	0.281
DYNL90	14.55	<0.001 *	4.65	0.032	0.08	0.777

Table 5. Analysis of variance summary table for the soft phonation contour at the successive frequency levels.

Note. df = 1, 222 for gender, vocal training, and gender x training. The figure behind DYNL indicates the frequency level, e.g., DYNL80 means the lowest phonation intensity at frequency level 80%. *p < 0.001

	Gender		Vocal training		Gender x training	
	F	p	F	p	F	p
DYNH0	0.29	0.591	7.89	0.005	8.14	0.005
DYNH10	18.84	<0.001*	12.91	<0.001*	1.42	0.235
DYNH20	32.78	<0.001*	6.81	0.010	0.03	0.874
DYNH30	23.38	<0.001*	1.17	0.280	0.55	0.461
DYNH40	7.54	0.007	0.82	0.368	1.35	0.247
DYNH50	0.94	0.333	0.38	0.538	0.07	0.787
DYNH60	8.08	0.005	2.68	0.103	1.32	0.252
DYNH70	1.70	0.194	9.05	0.003	0.80	0.371
DYNH80	16.27	<0.001*	4.48	0.036	1.89	0.171
DYNH90	13.52	<0.001*	0.35	0.557	0.12	0.726
DYNH100	8.60	0.004	0.06	0.800	4.13	0.043

Table 6. Analysis of variance summary table for the loud phonation contour at the successive frequency levels.
Note. df = 1, 222 for gender, vocal training, and gender x training. The figure behind DYNH indicates the frequency level, e.g., DYNH80 means the highest phonation intensity at frequency level 80%. *$p < 0.001$

Table 6 gives the results of the ANOVAs performed on the values of the loud intensities. No significant interaction between the factors gender and training was found. The factor gender showed a significant effect at the frequency levels 10, 20 and 30%, and 80 and 90%. At all these levels female subjects are able to phonate louder than male subjects. The factor vocal training showed an effect only at the frequency level 10%. At this level untrained subjects are able to phonate louder than trained subjects.

Conjoint frequency and intensity analysis

Applying the conjoint analysis method to phonetograms, the results obtained for female and male subjects are given in Table 7 and 8, respectively. Statistical analysis of differences in features was performed using Analysis of Variance (ANOVA) with factors gender and vocal training. T-tests were carried out at separate factor levels whenever significant interaction between the

Parameter	untrained (n=92) mean	SD	trained (n=42) mean	SD
Shape				
FD_1	0.11	0.064	0.12	0.078
FD_2	0.61	0.101	0.61	0.097
FD_3	0.21	0.093	0.16	0.089
FD_4	0.31	0.134	0.28	0.110
FD_5	0.17	0.102	0.17	0.094
FD_6	0.20	0.093	0.20	0.081
FD_7	0.17	0.096	0.15	0.082
FD_8	0.16	0.091	0.15	0.085
FD_9	0.15	0.091	0.15	0.084
FD_{10}	0.15	0.062	0.15	0.078
FD_{11}	0.11	0.057	0.12	0.066
FD_{12}	0.12	0.064	0.12	0.080
Contour Regularity	0.20	0.033	0.21	0.027
Slope (dB/st)	0.97	0.182	0.99	0.164
(dB/oct)	11.6	2.19	11.9	1.97
Area				
Enclosed area	0.21	0.049	0.28	0.058
Freq range (# oct)	2.93	0.385	3.32	0.399
Dyn. ranges and Central Position				
$WDR_{mff\text{-}3\ st}$	4.30	1.896	5.36	0.943
WDR_{mff}	5.31	1.146	6.02	0.365
$WDR_{mff+1/2\ oct}$	5.76	0.579	6.09	0.432
$WDR_{mff+1\ oct}$	5.40	0.975	5.89	0.489
$CP_{mff\text{-}3st}$	-0.98	0.895	-0.70	0.504
CP_{mff}	-0.33	0.606	-0.30	0.187
$CP_{mff+1/2\ oct}$	0.08	0.382	-0.04	0.158
$CP_{mff+1\ oct}$	0.33	0.533	0.12	0.223

Table 7. Analyzed characteristics of untrained and trained female phonetograms. Mean values and standard deviations (SD) are given. FD = Fourier Descriptor, WDR = Weighted Dynamic Range, CP = Central Position, mff = mean speaking fundamental frequency, st = semitones.

	untrained (n=47)		trained (n=43)	
Parameter	mean	SD	mean	SD
Shape				
FD_1	0.14	0.089	0.12	0.099
FD_2	0.57	0.089	0.54	0.090
FD_3	0.23	0.119	0.19	0.118
FD_4	0.25	0.107	0.24	0.128
FD_5	0.20	0.129	0.17	0.100
FD_6	0.24	0.101	0.21	0.114
FD_7	0.16	0.081	0.18	0.091
FD_8	0.18	0.094	0.15	0.092
FD_9	0.14	0.075	0.16	0.084
FD_{10}	0.16	0.098	0.14	0.071
FD_{11}	0.13	0.071	0.14	0.081
FD_{12}	0.12	0.075	0.14	0.067
Contour Regularity	0.21	0.034	0.22	0.032
Slope (dB/st)	0.91	0.208	0.94	0.191
(dB/oct)	10.9	2.49	11.3	2.30
Area				
Enclosed area	0.25	0.052	0.27	0.067
Freq range (# oct)	3.19	0.315	3.16	0.494
Dyn. ranges and central position				
$WDR_{mff-3\ st}$	3.37	1.843	4.59	1.504
WDR_{mff}	5.18	1.104	5.70	0.598
$WDR_{mff+1/2\ oct}$	6.01	0.408	6.21	0.247
$WDR_{mff+1\ oct}$	5.99	0.396	6.08	0.376
$CP_{mff-3st}$	-1.61	0.914	-1.06	0.734
CP_{mff}	-0.74	0.543	-0.49	0.284
$CP_{mff+1/2\ oct}$	-0.19	0.209	-0.10	0.121
$CP_{mff+1\ oct}$	0.10	0.193	0.18	0.174

Table 8. Analyzed characteristics of untrained and trained male phonetograms. Mean values and standard deviations (SD) are given. FD = Fourier Descriptor, WDR = Weighted Dynamic Range, CP = Central Position, mff = mean speaking fundamental frequency, st = semitones.

factors was found.

Shape. In all subgroups the amplitude of the Fourier Descriptors shows a pattern of alternating low and high values, with the high values diminishing and levelling off (see Tables 7 and 8). Using ANOVA inferential statistics the second Fourier Descriptor (FD_2) showed a significant difference between phonetograms of male and female subjects (see Table 9). The other descriptors are not significantly affected by these factors, however, with a probability level of 0.002, the third descriptor (FD_3) indicates a possible notable difference in averaged shape between untrained and trained subjects.

All values for contour regularity fall close together. The averaged contour regularity for male subjects is slightly higher with values of 0.21 for untrained and 0.22 for trained subjects. Female subjects exhibit values of respectively 0.20 and 0.21 for untrained and trained subjects. No significant effects of factors are found at a probability level $p<0.001$.

The slope of a line through the phonetogram with minimal distance to all phonetogram points is expressed both in decibels per semitone and decibels per octave in Tables 7 and 8. Generally, an increment of about one decibel per semitone is apparent in all subgroups. The average female slope values are slightly higher than the male, and trained subjects show higher values than untrained. However, none of these differences are significant.

Area. There is a large difference in averaged enclosed area between untrained and trained female groups. In the trained group almost three-tenths (0.28) of the reference area is covered by the phonetogram, compared with two-tenths (0.21) in the untrained female group (see Table 7). This difference is much smaller in male subjects, with values of 0.25 and 0.27 for untrained and trained, respectively (see Table 8). With ANOVA a significant effect of the factor vocal training could be established, trained subjects having a larger enclosed area (see Table 9).

In agreement with the frequency levels determined according to the rescaling method, the averaged frequency range in female subjects differs between untrained subjects (2.93 octaves) and trained subjects (3.32 octaves). The male subjects showed little difference, with an average frequency range of 3.19 for untrained and 3.16 for trained subjects.

ANOVA showed a significant interaction between the factors gender and vocal training, and therefore T-tests were performed at the separate factor levels. Within the female subgroup trained subjects showed a significantly larger frequency range compared with untrained subjects ($t = 5.46$, $df = 133$, $p < 0.001$). This difference could not be established in the male subgroup ($t = -0.31$, $df = 87$, $p = 0.761$). Within the untrained subgroup a significant difference in frequency range was observed between male and female subjects, with male subjects showing a larger frequency range ($t = 3.91$, $df = 137$, $p < 0.001$). In the trained subgroup no significant difference was present between male and female subjects ($t = 1.70$, $df = 83$, $p = 0.093$).

WDR and CP. The WDR, which is determined by adding the natural logarithms of the minimal and maximal intensity "distance" (in decibels) from the reference intensity of 75 dB, results in values for the intensity range in the speech area from about 4 to 6. From the results of the subgroups at the sampled frequency levels (Tables 7 and 8), the WDR is found to be maximal at the mean speaking F_0 plus half an octave ($WDR_{mff+1/2oct}$), and the trained groups have larger ranges

.

	Gender		Vocal training		Gender x training	
	F	p	F	p	F	p
Shape						
FD_1	1.42	0.235	0.01	0.923	1.36	0.245
FD_2	16.24	<0.001 *	1.00	0.318	1.83	0.178
FD_3	3.14	0.078	9.68	0.002	0.01	0.932
FD_4	10.70	0.001	2.19	0.141	0.58	0.447
FD_5	1.11	0.292	1.02	0.314	0.44	0.507
FD_6	4.19	0.042	0.99	0.321	1.15	0.285
FD_7	0.14	0.707	0.14	0.712	4.41	0.037
FD_8	0.70	0.403	1.83	0.178	0.39	0.535
FD_9	0.46	0.501	0.33	0.565	1.40	0.238
FD_{10}	0.12	0.725	0.53	0.467	2.65	0.105
FD_{11}	6.09	0.014	0.91	0.341	0.11	0.739
FD_{12}	0.42	0.520	0.44	0.508	0.62	0.432
Contour Regularity	4.16	0.043	4.35	0.038	0.16	0.686
Slope	3.50	0.063	0.72	0.397	0.00	0.968
Area						
Enclosed area	8.05	0.005	35.29	<0.001 *	11.55	0.001
Frequency range (# octaves)	2.30	0.131	14.47	<0.001 *	14.09	<0.001 *
"Speaking Range" Dynamics						
$WDR_{mff-3\ st}$	13.95	<0.001 *	23.57	<0.001 *	0.10	0.748
WDR_{mff}	2.51	0.114	22.89	<0.001 *	0.52	0.472
$WDR_{mff+1/2\ oct}$	9.53	0.002	17.94	<0.001 *	1.03	0.311
$WDR_{mff+1\ oct}$	18.46	<0.001 *	10.50	0.001	4.17	0.042
$CP_{mff-3\ st}$	16.50	<0.001 *	0.42	0.519	2.73	0.100
CP_{mff}	44.04	<0.001 *	0.38	0.541	4.59	0.033
$CP_{mff+1/2\ oct}$	22.76	<0.001 *	0.53	0.467	7.86	0.006
$CP_{mff+1\ oct}$	3.67	0.057	2.00	0.159	8.06	0.005

Table 9. Analysis of variance summary table for the variables of the conjoint method.
Note. df = 1, 222 for gender, vocal training, and gender x training.
FD = Fourier Descriptor; WDR = Weighted Dynamic Range; mff = mean fundamental speaking frequency; st = semitones; oct = octave; CP = Central Position. *p < 0.001 .

The calculated CPs of the WDRs show both negative and positive values, implying that the centre of the intensity range is below or above the reference intensity of 75 dB, respectively. The transition (with increasing frequency within the phonetogram) from a negative to a positive value occurs close to the sampled frequency mff+1/2 oct (see Tables 7 and 8).

ANOVA showed no significant interaction between the factors gender and vocal training with respect to the variables WDR and CP. The factor gender showed a significant effect on the WDRs at the mean speaking F_0 minus 3 semitones ($WDR_{mff-3st}$) and at the mean speaking F_0 plus an octave ($WDR_{mff+1oct}$). The female subjects have a larger range at the former sampled frequency, whereas the male subjects have a larger range at the latter frequency. Vocal training has a significant effect on the ranges at the first three sampled frequencies ($WDR_{mff-3st}$, WDR_{mff}, $WDR_{mff+1/2oct}$). Trained subjects have larger ranges at all these sampled frequencies.

The factor gender has a significant effect on the CP of the dynamic range at the first three sampled frequencies ($CP_{mff-3st}$, CP_{mff}, $CP_{mff+1/2oct}$). Male subjects have their CP of the dynamic ranges at a lower intensity value than female subjects at these sampled frequencies. No significant effect of the factor vocal training was found.

Comparison of phonetograms of untrained and trained subjects

Logistic regression was used to determine which variables best distinguish between untrained and trained phonetograms (21). With forward selection of most significant variables in an empirical model, logistic regression was applied to variables of phonetograms of both genders. Table 10 and 11 give a summary of significant variables that, in descending order, best distinguish between untrained and trained, for female and male subjects, respectively.

Variable	Significance	R
Enclosed area	0.0001	0.27
$WDR_{mff+1/2\ oct}$	0.0001	-0.28
WDR_{mff}	0.0002	0.26

Table 10. Logistic regression applied to the factor training in female subjects. Most significant variables are listed with significance level and correlation coefficient. The model is evaluated with a goodness of fit and percentage of correctly classified subjects. Goodness of Fit 0.8261 Correctly classified 83.7% (+14.8%)

The variable that best distinguishes between untrained and trained phonetograms of female subjects turned out to be the enclosed area, followed in order by the Weighted Dynamic Ranges at the mean speaking fundamental frequency + 1/2 octave ($WDR_{mff+1/2oct}$) and at the mff (WDR_{mff}), respectively. A fourth significant variable is the Central Position of the WDR at mff + 1/2 octave ($CP_{mff+1/2oct}$). The level of significance and correlation coefficient are given in Table 10. The percentage of correctly classified subjects with introduction of all four significant variables is 83.7%, which adds 14.8% to the coincidentally correctly classified

Variable	Significance	R
$WDR_{mff\text{-}3\,st}$	0.0387	0.14
FD_2	0.0069	-0.21
$CP_{mff+1\,oct}$	0.0010	0.27
$WDR_{mff+1/2\,oct}$	0.0367	0.14
FD_7	0.0524	0.12

Table 11. Logistic regression applied to the factor training in male subjects.
Goodness of Fit .6284
Correctly classified 75.3% (+23.6%)

percentage. The goodness of fit, a measure for the overall explanatory power of the procedure, is 0.83 of a maximum value of 1.0 .

In male subjects the lowest WDR ($WDR_{mff\text{-}3st}$) is the variable that is first selected in distinguishing between untrained and trained, followed by the second Fourier Descriptor (FD_2), the Central Position of the WDR at mff + 1 oct($CP_{mff+1oct}$), the WDR at mff + 1/2 oct ($WDR_{mff+1/2oct}$)and finally the seventh Fourier Descriptor (FD_7). Table 11 is also completed with significance and correlation coefficient values of each variable present in the empirical model. The value of 0.63 for the goodness of fit is slightly lower than the female value, but a large percentage of correctly classified subjects can be found (75.3%, which adds 23.6%).

DISCUSSION

Rescaling method

Analysis of phonetograms with the rescaling method offers results that can be compared with data of previous studies.

Schultz- Coulon (16) acquired phonetograms of 21 female and 25 male students without vocal training, and presented summarizing statistics of these phonetograms as "Normstimmfeld" (norm phonetogram) for the two groups. When these "Normstimmfelde" are compared to the untrained phonetograms of Figures 2 and 3, similarities as well as a few differences are apparent. The lowest phonatory frequency is comparable in male phonetograms, 87.3 and 86.1 Hz in the Schultz-Coulon and the present study, respectively. In the female "Normstimmfeld" this frequency level is 10 Hz lower (147 versus 157.3 Hz, ca. 1 semitone). The local minimum in the loud phonation contour, found in the averaged untrained female and male phonetogram, is not present in the "Normstimmfelde". The minimum intensities in the present study are a few decibels lower (about 1 dB for the averaged female phonetogram and about 3 dB for the averaged male phonetogram). The exact differences cannot be presented, as the Schultz-Coulon study does not give a numerical representation of the mean values. The small difference in minimum intensities might be explained by the use of a dB(A) weighting network in the present study (7). Mean maximal dynamic ranges in the present study for untrained subjects are 39.1 and 42.8 dB for female and male phonetograms, respectively, both at the 40% frequency level. These ranges are 5.5 and 3.7 dB smaller than the ranges given by Schultz-Coulon.

Awan (17) gives tabulated results of the mean extreme intensities of phonetograms. However, in the Awan study gender is not separately treated while discussing mean phonetograms, which limits the possibilities for comparing results. Another difference with the present study is the requirement of "quality" phonations in registering a phonetogram. Therefore it is not surprising that the untrained subjects in the present study produced louder phonations (about 15 dB), and, as phonations with audible breathiness were accepted, the minimum phonations are softer in the modal register. Differences between the studies are much smaller when comparing the maximum (loud) phonations for trained subjects: the present study gives phonations that are only a few decibels lower at the lower frequency levels and a few decibels higher at the centre frequencies. However, the minimum (soft) intensities are much softer (ca. 20 dB), a difference that can probably be attributed to the use of "quality" phonations in the Awan study. The fact that specific differences in mean intensity levels between male and female subjects have been ascertained is an indication that male and female results might preferably be discussed separately in future studies.

Åkerlund (18) compared phonetograms of female singers and non-singers with the rescaling method. Results showed that singers had the ability to phonate at slightly lower intensities almost over the entire frequency range. This corresponds with the finding in the present study. However, the absolute minimum intensities are higher in the Åkerlund study and the slope of the soft phonation contour is less steep. Both differences can partially be explained by the use of the dB(A) weighting network. The non-singers in the Åkerlund study demonstrate limited loud phonation capacities, where the singers' superior abilities to produce sound in the falsetto register are reflected in the intensity levels at higher frequencies. In the present study a smaller difference between the two groups was observed in this frequency range, and absolute intensities measured fall between the singers and non-singers of the Åkerlund study. The higher intensities of Åkerlund's singers might be related to their superior training, which took place in a musical conservatory over at least four years.

Though not significant with the rescaling method, the louder sound intensities in male singers (compare Figure 3 above and below) could well be related to a laryngeal disposition enabling flow phonation in these persons (22). Also, a better control of tuning the spatial dimensions of the vocal tract enhances the resonance effect, thereby increasing the sound intensity (23).

Conjoint analysis

Shape. Fourier Descriptors. Fourier Descriptors (FDs) offer the possibility to give a numerical representation of a shape (19). Having quantified the complex appearance of a phonetogram, this information can be used to compare with other analyzed phonetograms or to refer to the general shape of a 'normal' phonetogram. Deviations from mean FD values, which imply departures from normal shape, can also be quantified.

To ensure a meaningful use of the analytic power of FDs, attention must be paid to phonetographical procedures. As the shape is influenced by the number of samples along the individual frequency range, we recommend the use of a specific number of samples per octave. In this study a sampling frequency of four samples per octave was used.

On the one hand, a certain minimum of FDs are needed to give an acceptable representation of a phonetogram; on the other hand, measuring error in intensity, causing only small changes in the shape of a phonetogram, is mainly reflected in the higher order FDs. With these considerations in mind we settled on the 12 FDs.

According to the values of the FDs the differences in general shape between untrained and trained are small: only the third FD (FD_3) has a low probability level ($p = 0.002$). These overall minor differences in FD values are consistent with the gross equality in shape of the averaged phonetograms. The most notable exception is the local minimum in the loud phonation contour at the frequency level where the transition from modal to falsetto register occurs. The gradual transition from modal to falsetto register can probably be attributed to a better control over laryngeal musculature in the trained group (24).

Contour regularity. The larger contour regularity in the trained group indicates a relatively shorter perimeter of the phonetogram with respect to the enclosed area. This means either that there is less irregularity in the perimeter of the phonetogram, or the phonetogram has a more circular shape (since a circle produces the maximal value for contour regularity). The absence of the local minimum in the loud phonation contour of the trained phonetogram reduces the perimeter, thereby increasing the value of the contour regularity. A second consideration regarding contour regularity is that an untrained voice might experience greater difficulties in producing phonations at the physiologic boundaries, causing an irregular aspect of the perimeter. These two observations give an explanation for the larger contour regularity in the trained voice.

Slope. In a number of articles an indication is given about the slope of phonetograms. However, comparing slopes among studies is difficult, because of the different ways of defining this variable.

Treating registers independently, Klingholz (25) analyzed slopes of phonetograms of untrained and trained subjects. Values varied between 3.0 and 2.1 dB/st for untrained and trained female subjects, respectively, in chest register and 4.0 and 3.3 dB/st for the male counterpart. The slope of the falsetto register, however, shows a different aspect. These values vary between 0.5 dB/st for both female groups and 0.6 and 0.4 for the untrained and trained male phonetograms, respectively. The values for slopes in this study with 0.97 and 0.99 dB/st for untrained and trained female subjects, respectively, and 0.91 and 0.94 dB/st for untrained and trained male subjects, respectively, fall in between the values presented from the Klingholz study and therefore are not necessarily in contradiction.

Awan (17) gives a separate slope for maximum and minimum intensity after rescaling phonetograms. Transposing the given slope, based on an x-axis with 10% frequency levels, to a slope with dB/st units for the maximum SPL this yields values of 0.36 and 0.67 for untrained and trained subjects, and for the minimum SPL values of 0.51 and 0.78 are calculated. These values differ from the mean slopes found in this study. An explanation can perhaps be found in the exclusive acceptance of "quality" vowels in the Awan study, as well as the use of a dB(A) weighting scale in the present study.

The theoretically derived slope of 8-9 dB/octave in the study of Titze (8) for the soft phonation contour is close to the values found in this study. In another study with empirical data in a physiologic model, Titze (26) arrives at the same rise in intensity of 8-9 dB/octave. This slope was determined assuming that the lung pressure remained proportional to the phonation threshold pressure (see also Titze 22).

Finally, results of a study by Sundberg et al.(27) showed that if, over an octave, the excess subglottal pressure over threshold pressure is doubled, the resulting intensity slope is 8-9 dB/octave.

Area. Enclosed area. The trained groups have larger enclosed areas. Of all groups, the female trained subjects have the largest value, which means that their combined capacities in the frequency and dynamic range are unsurpassed. The experience of trained sopranos in using the falsetto register makes a major contribution to this overall superiority.

Frequency range. Almost all articles on vocal capacities report a mean frequency range for a group of subjects. Awan (17) gives a summary of these F_0 range studies: values for frequency range run from 2.60 (25) to 3.16 octaves (28), and from 2.73 (29) to 3.12 (6) for untrained male and female subjects, respectively. Klingholz (25), however, reports a much lower frequency range of 2.38 octaves for a group of untrained female subjects, and Awan (17), using a "musical" F_0 range, gives a range of 2.29 octaves for a combined group of female and male subjects.

In the present study ranges of 3.19 and 2.93 octaves were found for groups of untrained male and female subjects, respectively. The male range is thus slightly above ranges previously reported.

Limited information is available on the frequency ranges of subjects with vocal training. A number of studies present a range of about 2.9 octaves for trained male subjects (17,25,30) and about 3.0 octaves for female subjects (17,25), whereas in this study values of 3.16 and 3.32 octaves are found. Thus both ranges are a few semitones higher than the previously reported values. Though non-significant, the less extensive frequency range in trained male subjects might be attributed to two factors. The first of these is that, while a number of basses and baritones have superior control in the low frequency range, their capacities are limited at high frequencies. The second factor, for which we offer no explanation, was that a number of trained male subjects were not able to produce phonations in falsetto register.

WDR and CP. WDR. The highly significant differences in WDRs are evidence of the superior capacities of trained subjects in varying sound intensity. In trained male subjects these large dynamic ranges in the low frequency region are often required for an optimal singing performance as a bass, baritone or tenor. Female trained subjects also seem to benefit from the experience of varied exploitation of the voice source, which results in an enhanced control of the intensity output.

CP. The CP at a lower intensity level indicates that the main part of the extended dynamic ranges characteristic of trained subjects is located in the soft speaking voice area. Female trained subjects, who also have a relatively extended soft speaking voice area at the lowest sampled frequency, show in addition an extended loud speaking voice area at the highest sampled frequencies. Subjects with limited control over the voice source show inferior phonation capacities in this frequency region.

Differences between untrained and trained For both genders logistic regression is the analytic procedure that reveals the most significant differences between untrained and trained phonetograms.

In female subjects the enclosed area, comprising capacities in frequency range as well as dynamic ranges, is the main difference between the two groups. The other significant variables, mainly consisting of WDRs at frequency levels in the centre of the modal register, reveal less striking differences.

In male subjects the large dynamic capacities of the trained group at low frequency levels is apparent. In this frequency region the group of basses and baritones demonstrate their superior control. Also some shape features, expressed in Fourier Descriptors, are significantly different. Topics for further investigation might include these questions: which aspects of shape are represented by these Fourier Descriptors, and how might other subsets of subjects, for instance specific groups of patients, show a characteristic departure from the normal pattern of FD values?

CONCLUSIONS

Results of this study indicate differences between male and female, as well as between untrained and trained phonetograms. These differences can be attributed to greater natural capacities in trained subjects or a superior learned control over the voice mechanism. The anatomical constitution of the larynx provides varying vocal capacities among persons. In exploring the physiologic boundaries of these capacities, the ability to control the voice source is directly related to the measured dynamic and frequency ranges. The singing experience is associated with improved neuromyogenic control over the voice source.

Male subjects were able to phonate more softly than female subjects, whereas female subjects could phonate louder at distinct low and high relative frequency ranges (10-30% and 80-90%). The dB(A) filter could explain these differences in intensity production for the lowest frequencies, but not at the frequency levels above 40%. A difference in shape of the averaged male and female phonetogram was detected by a characteristic difference in one of the Fourier Descriptors, indicating the potential analytic power of the shape-quantifying procedure.

Singers, especially the male group, demonstrate their ability to produce vocal sound at low frequencies, resulting in a phonetogram with larger dynamic capacities at these frequencies. Both female and male singers have larger dynamic ranges, especially in soft voice. The enhanced possibilities with the soft voice suggest a better breath control during voicing and the ability to oscillate vocal folds at lower subglottal pressures. At the ends of the frequency range untrained subjects produced louder intensities. However, these differences are a result of the rescaling procedure, which makes comparisons at relative frequency levels. Had the comparisons been made at absolute frequencies, the untrained subjects would have shown no such louder intensities.

Another common feature of trained phonetograms is the relative absence of a local minimum at the frequency range where the register transition takes place. This study, with a large number of subjects, confirms the preliminary conclusions of other studies.

In this investigation we looked at differences between untrained and trained groups, however, the data of the untrained group might also be used in a similar way to detect specific differences in phonetograms pertaining to particular pathologies or as a reference database for comparing individual phonetograms.

One should be careful not to attribute the superior phonetograms of the "trained" group simply to the effects of training. In the first place, what distinguished this group was not training as such, but, in most cases, participation in choral singing. In the second place, better capabilities of the voice source may have led these people to singing, rather than resulting from the singing activity. The results of this study, however, suggest a comparable constitution of the voice source in trained and untrained subjects, while the extended phonatory capabilities in trained subjects are probably based on an improved voluntary control over the voicing mechanisms. A forthcoming article about a study on vocal physiology in the same specified groups, as well as a prospective study analyzing the effect of vocal training, will address these hypotheses.

References

1. Wolf S, Stanley D, Sette W. Quantitative studies on the singing voice. J Acoust Soc Am 1935;6:255-66.
2. Damsté P. The phonetogram. Pract Otorhinol 1970;32:185-7.
3. Schutte H. Over het fonetogram. Logopedie en Foniatrie 1975;47:82-92.
4. Schutte H. Untersuchungen von Stimmqualitäten durch Phonetographie. HNO Praxis (Leipzig) 1980;5:132-9.
5. Schutte H. Belastbaarheid van de stem en het fonetogram. Logopedie en Foniatrie 1986;58:140-3.
6. Coleman R, Mabis J, Hinson J. Fundamental frequency-sound pressure level profiles of adult male and female voices. J Speech Hear Dis 1977;20:197-204.
7. Gramming P, Sundberg J. Spectrum factors relevant to phonetogram measurement. J Acoust Soc Am 1988;83:2352-60.
8. Titze I. Acoustic interpretation of the voice range profile (phonetogram). J Speech Hear Res 1992;35:21-34.
9. Sulter A, Wit H, Schutte H, Miller D. A structured approach to Voice Range Profile (phonetogram) analysis. J Speech Hear Res 1994;37;1076-85.
10. Stell P, Gregory I, Watt J. Morphometry of the epithelial lining of the human larynx. I.The glottis. Clin Otolaryngol 1978;3:13-20.
11. Holmberg E, Hillman R, Perkell J. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. J Acoust Soc Am 1988;84:511-29.
12. Schutte H. The efficiency of voice production. San Diego: Singular Publishing Group, 1980
13. Shipp T, Izdebski K. Vocal frequency and vertical larynx position by singers and nonsingers. J Acoust Soc Am 1975;58:1104-6.
14. Watson P, Hixon T. Respiratory kinematics in classical (opera) singers. J Speech Hear Res 1985;28:104-21.
15. Schutte H, Seidner W. Recommendation by the Union of European Phoniatricians (UEP): Standardizing Voice Area Measurement/Phonetography. Folia Phoniatr 1983;35:286-88.
16. Schultz-Coulon H, Asche S. Das "Normstimmfeld" - ein Vorschlag. Sprache-Stimme-Gehör 1988;12:5-8.
17. Awan S. Phonetographic profiles and F0-SPL characteristics of untrained versus trained vocal groups. J Voice 1991;5:41-50.
18. Åkerlund L, Gramming P, Sundberg J. Phonetogram and averages of Sound Pressure Levels and fundamental frequencies of speech: Comparison between female singers and nonsingers. J Voice 1992;6:55-63.

19. Zahn C, Roskies R. Fourier Descriptors for plane closed curves. IEEE Trans Comp 1972;21:269-81.
20. Coleman R. Sources of variation in phonetograms. J Voice 1993;7:1-14.
21. Hair J, Anderson R, Tatham R, Black W. Multivariate data analysis. New York: MacMillan Publishing Company, 1992.
22. Titze I, Sundberg J. Vocal intensity in speakers and singers. J Acoust Soc Am 1992;91:2936-46.
23. Miller D, Schutte H. Formant tuning in a professional baritone. J Voice 1990;4:231-7.
24. Murray T, Caligiuri M. Phonatory and nonphonatory motor control in singers. J Voice 1989;3:257-63.
25. Klingholz F, Martin F. Die quantitative Auswertung der Stimmfeldmessung. Sprache-Stimme-Gehör 1983;7:106-10.
26. Titze I. Phonation threshold pressure: A missing link in glottal aerodynamics. J Acoust Soc Am 1992;91:2926-35.
27. Sundberg J, Titze I, Scherer R. Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. J Voice 1993;7:15-29.
28. Hollien H, Dew D, Philips P. Phonational frequency ranges of adults. J Speech Hear Res 1971;14:755-60.
29. Ptacek P, Sander E, Malone W, Jackson C. Phonation and related changes with advanced age. J Speech Hear Res 1966;9:353-60
30. Colton R, Holllien H. Phonational range in the modal and falsetto register. J Speech Hear Res 1972;15:708-13.

9

Effects of voice training on phonetograms and maximum phonation times in female speech therapy students

Arend M. Sulter and Jorien M. Meijer

Department of Otorhinolaryngology, University Hospital Groningen, Groningen, the Netherlands

Submitted

INTRODUCTION

Voice training affects morphological aspects of, as well as the control over the voice source. With specific instructions, benign lesions of the vocal folds, specifically vocal fold nodules (1-3), decrease in size (4-6), and in some cases the closure of vocal folds can be improved (7, 8). The instructions given during voice therapy or training modify and improve laryngeal muscle strength, tone, balance, and stamina (9). By establishing conditions for a healthy vocal fold cover, these instructions result in ameliorating the symptoms of benign lesions and in many cases prevent a recurrence (10-12). As trained subjects, singers also elicit effects of the improved control over the voice source. A number of studies exemplify the differences between singers and nonsingers with respect to motor control (see Murry & Caligiuri) (13).

Improved laryngeal muscle strength and tone, as well as improved balance among laryngeal muscle effort, respiratory effort and control, might result in an increase in vocal capacities. These capacities can be visualized in a phonetogram (14). Figure 1 gives an example of a phonetogram with lines representing frequency and intensity ranges. Comparing phonetograms of

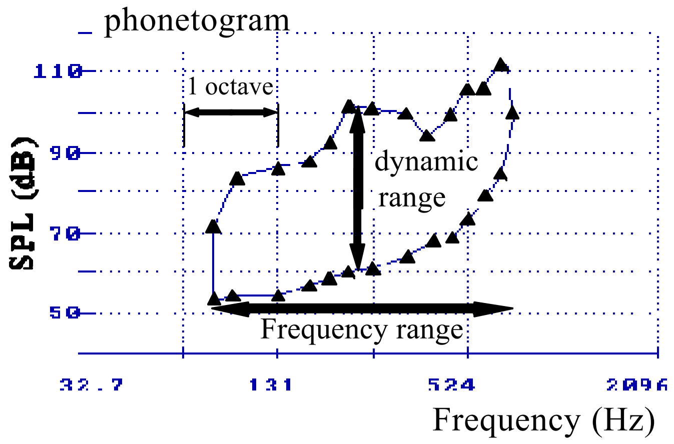


Figure 1. An example of a "normal" male phonetogram. Along the x-axis the frequency scale is plotted (32.7 - 2096 Hz) and the intensity level is given along the y-axis (40 - 120 dB). Note the dip in the loud phonation contour at about 400 Hz. This local minimum exhibits the transition of chest register to falsetto register. The lines in the vertical and horizontal direction indicate the dynamic and frequency range, respectively.

singers and nonsingers, the singers show greater dynamic capacities (15) and extended frequency ranges (16, 17). In a study comparing untrained and trained vocal groups, Awan (18) found a significant correlation between speaking range dynamics and both intensity and frequency ranges of phonetograms. Therefore a change in phonetogram profile is expected after voice training, reflecting changed speaking range dynamics. However, with speech training only increased intensities are to be expected, as phonation with fundamental frequencies beyond pitches used in normal speech is not practiced.

Another method to analyze voice function is measuring maximum phonation times (19). The s/z ratio has previously been employed as an indicator of laryngeal pathology (20). As the maximum phonation time of the /s/ reflects the expiratory flow control and the /z/ indicates glottal resistance, the s/z ratio might change after voice training.

Most previous investigations of voice training effects are cross-sectional studies. A possible a priori difference between groups, independent of voice training, therefore can not be excluded. A prospective study is more suitable for analyzing effects of voice training.

The specific content of voice training depends on the intended goal. Training given to voice patients is directed toward curing or ameliorating pathology, correcting pathogenic voicing patterns, as well as optimizing the use of voice capacities. The optimization is achieved by correcting breath control, promoting relaxed phonation and employing resonance potencies by adjusting vocal tract anatomy.

Singing differs from speech with respect to intensity and frequency ranges, breath management and neuromyogenic control (21). Voice training given to singers focuses on improving these aspects of phonation and therefore differs from speech training.

In this study effects of specific speech training given to female speech therapy students are analyzed by comparing phonetograms and maximum phonation times before (PRE) and after (POST) two and a half years of training.

METHODS

Subjects

A group of 25 female students (age 17 - 21 years, mean age 18.4 years, standard deviation (SD) 1.20 years) was investigated before starting the study of speech therapy at the Training School for Speech Therapist, Groningen. After two and one half years of education the same investigation, among others consisting of phonetography and measuring phonation times, was repeated. Pathology of vocal folds was excluded by laryngostroboscopic examination of the students. The effect of ageing of the students on the measured variables was considered by analyzing a data base with normative data on a comparable age cohort (17).

Voice training

Voice training entailed practicing several methods during speech therapy study in Groningen. The goal of voice training was to attain a clear voice carriage with adequate (normal tonus) bodily posture and diaphragmatic breathing. The training methods are described in short beneath. For a more detailed description of the exercises the reader is referred to the literature (see references).

The exercises of Coblenzer (22, 23) required coupling breathing, phonation and rhythmical movements. Essential in this method is the bodily posture (eutonus), comprising the balance among laryngeal muscle effort, respiratory effort and control, and supraglottal modification of the laryngeal tone. This posture enables maximum vocal performances with a minimum of effort. When the right balance in tension has been found, the attention is directed toward breath management. The last part aims at learning to use reflectory movements of the diaphragm to provide automatic inhalation.

The resonance method (24, 25) required a proper balance between stress and relaxation during phonation. Manipulating resonance characteristics by adjusting vocal tract configuration in combination with a low pharyngeal position, the voice gains intensity with clear carriage.

The Smith Accent Method (26) aims at the development of abdominaldiaphragmatic breathing using accentuated rhythmic movements. This facilitates optimal breath management, resulting in a firm glottal closure. The gist of the method is practicing in chest register with vocal folds that are short and lax, and therefore have a large vibrating mass.

The nasalizing method (27) derives its name from a striking and characteristic part of the method: the nasalization of sound. The nasal sound originates from a completely relaxed larynx and vocal tract, where the air flows both through mouth and nose. The soft palate hangs downward loosely, which results in a relaxed "lifting" mechanism. The relaxation has a beneficial effect upon voice function. The resonant space is larger because of a wider and longer vocal tract, and the resonant quality is improved by the relative absence of tension in the larynx and vocal tract. The possibilities for modulating articulations are improved in combination with less eminent vocal and speech fatigue.

The students were all equally trained during a period of two and a half years. The forementioned methods were exercised and practiced for a total of 260 hours of education per student.

Phonetography

Equipment. Phonetograms were registered in a sound treated room with "living room acoustics" (28), using an FST-II phonetometer. This phonetometer and its operation are described in detail in Sulter et al. (17).

All phonetograms were registered by two investigators, both familiar with the equipment and phonetographic procedure, and having had more than five years musical training and experience. Both investigators used a standard set of detailed instructions for the students.

Phonetographic Procedure. Phonetography recommendations by the UEP were followed (28). The direction and distance (30 cm) of the student's mouth to the microphone was carefully controlled during the procedure. The students were tested using the vowel /a/. Phonetogram points were collected, starting the acquisition at the mean speaking fundamental frequency, followed by the low frequencies and ending with high frequencies. The mean speaking fundamental frequency was determined by asking the students to count from one to ten. Then the investigator's imitation of that frequency was measured with the phonetometer. A reproducible phonation at a given frequency with a minimum phonation time of one second, to insure a stabilized sound intensity production and correct measurement, was required for accepting a phonetogram point. During actual registration, the student, if necessary, was at first guided in matching the target frequency, and after this the minimum and maximum intensity were registered. The students were instructed to produce phonations at the physiologic boundaries, without, of course, injuring the voice during maximum intensity. The frequency range was sampled at four frequencies per octave, basically at the tones c-e-g-a, e.g., C4-E4-G4-A4 (in this octave, 262, 330, 392 and 440 Hz). At the upper and lower ends of the range, shorter frequency intervals of semitones were chosen. The phonetograms were acquired within a ten to twenty minute time period.

Phonetogram Analysis. For reasons of comparison and establishing differences between phonatory capacities a standardized approach was used. Phonetograms were analyzed using two different methods, which are described in detail in Sulter et al. (17, 29). What follows is a brief description of these methods.

Rescaling method. The rescaling method determines dynamic ranges at fixed relative distances along a student's (individually differing) frequency range (see Figure 1). Retrieving phonetogram points from individual data files, each frequency range was rescaled to 100% with a specially developed computer program. At 10% intervals the minimum and maximum intensity was calculated by interpolation, yielding eleven values for both minimum and maximum sound pressure levels. By averaging the SPL values at each interval for a number of students, normative data on the dynamic ranges were established.

Conjoint frequency and intensity analysis of the phonetogram. Another approach for the analysis of phonetograms concentrates on the phonetogram features shape, area and Weighted Dynamic Range and Central Position. The main difference between this approach and the rescaling method is that it derives variables without distortion of the shape.

The shape of a phonetogram is described with so-called Fourier Descriptors (FDs), thus enabling quantification (29, 30). Changes in shape can be notified with these FDs. Shape can also be described by contour regularity, which is the quotient of enclosed area and squared perimeter, producing a dimensionless value. The slope of a phonetogram is determined by drawing a line with minimal distance to all points through the phonetogram and calculating the angle between this line and x-axis.

The enclosed area gives the quotient of the voice field area and the rectangle with coordinates 40 and 110 dB, and 32.7 and 2096 Hz. The frequency range is derived by subtracting the lowest phonated fundamental frequency from the highest frequency.

A Weighted Dynamic Range (WDR) is determined in the modal register at four frequencies[1] which are related to the mean speaking fundamental

---

[1]The four sampled frequencies in the modal register are: mean speaking fundamental frequency (mff) minus three semitones; mff; mff plus half an octave; and mff plus an octave. In this study the mff was standardized at 220

frequency. Because sound intensities around 75 dB are relatively more important in normal speech than very soft and loud intensities, a logarithmic weighting factor was used, giving greater weight to the intensities in accordance to how close they are to 75 dB. Besides the WDR, the Central Position (CP) of this range is calculated. For a detailed description the reader is referred to Sulter et al. (29).

Both these methods for analyzing phonetograms were implemented in a computer program written in the ASYST language (ASYST, Macmillan software company) and running in DOS environment.

Maximum phonation times

After a demonstration by the examiner, the students were instructed to inhale and to produce the vocalized consonant /z/ for as long as possible at a comfortable pitch and loudness level. The same instructions were given for the production of a sustained /s/. Each student produced the consonants twice. Productions were measured with a stopwatch with an accuracy of 0.1 s. The longest phonation of each consonant was used for further analyses.

Statistical analysis

To analyze the effect of voice training, comparisons were made between PRE and POST phonetograms, as well as between maximum phonation times. Differences in variables resulting from the rescaling method were analyzed with two-way analysis of variance (ANOVA) with rescaled frequency value and voice training (PRE/POST) as factors. To detect significant differences among factor levels, post hoc Tukey HSD tests were performed. Differences between variables resulting from the phonetographical conjoint analysis method, as well as differences in maximum phonation times were analyzed with paired t-tests. Linear regression was used to analyze the influence of age on phonetographical variables and maximum phonation times. A probability level of 0.05 was used to reject the null hypothesis, which posits no difference among variables under investigation.

RESULTS

First the results of comparing phonetograms will be given, followed by a presentation of the results of maximum phonation times.

Phonetograms

Rescaling Method. After averaging the 25 phonetograms with the rescaling

Hz.

method, descriptive statistics can be derived for each rescaled frequency value and vocal intensity. These average rescaled frequency values will first be discussed, and then the average minimum and maximum intensity levels will be presented. Inferential statistics are used to establish differences in untrained and trained phonetograms.

Rescaled frequency values. Table 1 gives average frequencies with standard deviations at the 10% rescaled frequency values.

An average frequency range from 158.4 Hz to 1255.7 Hz was measured in the untrained female students (PRE). After two and a half years of education (POST) average frequency values ranging from 157.9 to 1212.5 were established. Thus the trained students have a reduced frequency range and phonate slightly lower at the 100% rescaled frequency value; however, this difference is less than one semitone.

Frequency level		PRE	POST
0%	mean (Hz)	158.4	157.9
	SD (Hz)	23.78	19.01
10%	mean (Hz)	194.2	193.2
	SD (Hz)	25.77	20.83
20%	mean (Hz)	238.2	236.4
	SD (Hz)	28.49	23.82
30%	mean (Hz)	292.6	289.4
	SD (Hz)	32.69	29.01
40%	mean (Hz)	359.5	354.4
	SD (Hz)	39.85	37.66
50%	mean (Hz)	442.0	434.3
	SD (Hz)	51.76	51.14
60%	mean (Hz)	543.7	532.8
	SD (Hz)	70.26	70.59
70%	mean (Hz)	669.6	653.6
	SD (Hz)	98.08	98.31
80%	mean (Hz)	825.1	802.7
	SD (Hz)	137.58	136.38
90%	mean (Hz)	1017.5	986.2
	SD (Hz)	192.96	188.13
100 %	mean (Hz)	1255.7	1212.5
	SD (Hz)	269.10	257.12

Table 1. Mean frequencies and standard deviations (SD) for consecutive frequency levels of a group of speech therapy students at the beginning of the study (PRE) and after two and a half years of education (POST).

The standard deviation shows an absolute increase in frequency as the rescaled value rises. In semitones this amounts to about 2 semitones at the 0% rescaled frequency values and 3 semitones at the 100% rescaled frequency values.

Sound intensity levels. Table 2 gives the average maximum and minimum phonation intensities, together with standard deviations at the 10% rescaled frequency values. Using the average frequencies of Table 1, Figure 2 was constructed, visualizing the average phonetograms for untrained (PRE) and trained (POST) students.

The standard deviations for frequencies and intensities are represented as whiskers in both plots. Apart from the values near the extreme rescaled frequency values, the standard deviations of intensities are generally less than 6 dB. With increasing frequency a rising intensity level is present for the soft phonation contour in both the untrained and trained average

Frequency level		Maximum PRE	Maximum POST	Minimum PRE	Minimum POST
0%	mean (dB)	66.1	69.4	56.4	51.7
	SD (dB)	10.80	9.10	7.38	6.57
10%	mean (dB)	86.2	87.6	54.8	47.9
	SD (dB)	4.91	5.33	6.40	3.09
20%	mean (dB)	91.8	95.4	57.3	50.1
	SD (dB)	3.86	3.55	5.77	3.60
30%	mean (dB)	96.5	98.7	59.3	52.3
	SD (dB)	5.08	2.54	5.78	3.22
40%	mean (dB)	99.2	101.1	61.2	54.7
	SD (dB)	5.47	4.73	6.10	3.25
50%	mean (dB)	99.3	103.3	64.4	57.3
	SD (dB)	6.23	4.92	7.33	3.94
60%	mean (dB)	99.1	104.1	68.7	60.8
	SD (dB)	5.01	5.44	7.60	5.66
70%	mean (dB)	101.2	103.3	74.0	65.5
	SD (dB)	5.14	3.12	7.14	6.34
80%	mean (dB)	104.1	104.4	80.1	72.6
	SD (dB)	6.05	4.05	7.84	9.08
90%	mean (dB)	104.9	106.3	87.0	79.8
	SD (dB)	7.15	5.43	8.76	9.56

Table 2. Maximum and minimum intensities in a group of speech therapy students at the beginning of the study (PRE) and after two and a half years of education (POST). Mean intensity levels and standard deviations (SD) are given.

phonetogram. The average trained (POST) phonetogram shows better possibilities for producing soft intensities, compared to the average untrained phonetogram.

A distinct difference in the loud phonation contour between the average phonetograms is the location of a relative minimum in an otherwise continuously rising loud phonation contour, which is present at the 60% rescaled frequency value (about 500 Hz) of the average untrained (PRE) phonetogram, whereas it is located at the 70% rescaled frequency value (about 600 Hz) in the average trained (POST) phonetogram. This relative minimum marks the transition from modal to falsetto register, which therefore seems to occur at a higher frequency in the trained group.

Table 3 gives the results of the ANOVA procedure applied to frequency and intensities. No significant interaction was found between the factors rescaled frequency value and training. Not surprisingly, the effect of the factor rescaled frequency value on frequency is highly significant. The factor vocal training (PRE/POST) has no significant effect on calculated frequencies. Both the factor rescaled frequency value and vocal training have significant effects on maximum and minimum intensity.

Post hoc Tukey-HSD tests with a significance level $p=0.05$ performed for the factor vocal training showed a significant difference in maximum intensity

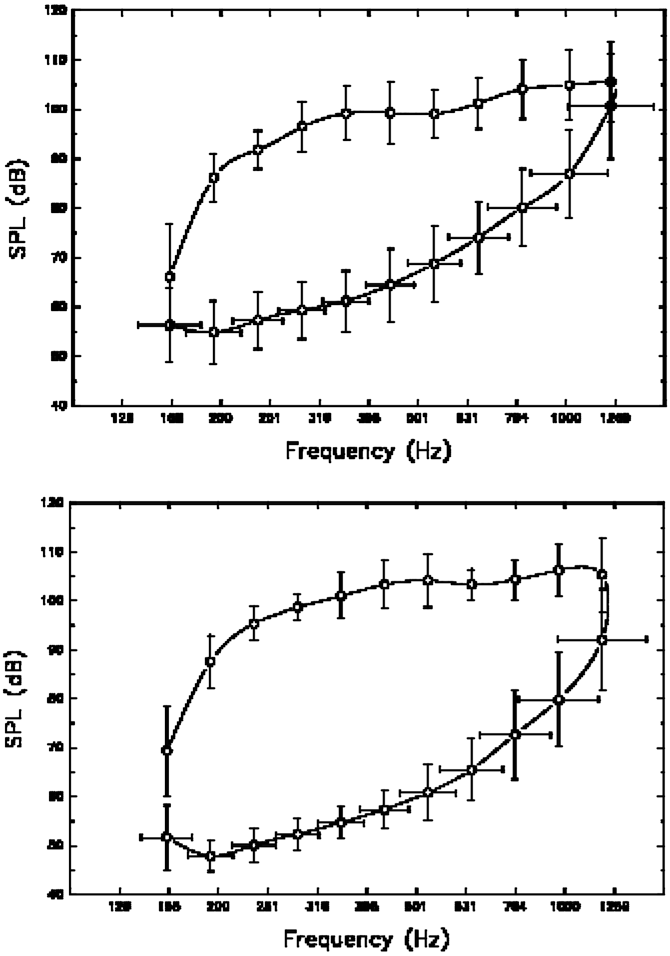


Figure 2. Averaging frequency and intensity data, mean phonetograms are plotted for a group of 25 female voice therapy students (above) at the beginning of the study and (below) after two and a half years of education. The whiskers indicate 1 standard deviation.

at the 20%, 50% and 60 % rescaled frequency value. At all rescaled frequency values significant differences in minimum intensities were found.

Conjoint Frequency and Intensity Analysis. The results obtained with the conjoint analysis method are given in Table 4. Paired t-tests were performed to detect significant differences between average variables from untrained and trained phonetograms.

Shape. Before and after the period of vocal training the average shape shows Fourier Descriptors with the previously described pattern of alternating low and high values (17). For the higher order FDs the high values diminish and level off (see Table 4). A significant difference in average shape was established for the first (FD_1), second (FD_2), as well as the seventh

	Frequency level		Voice training	
	F	p	F	p
Frequency	1162.9	<0.001 *	2.1	0.149
Maximum intensity	177.0	<0.001 *	20.2	<0.001 *
Minimum intensity	211.0	<0.001 *	146.9	<0.001 *

Table 3. Analysis of variance summary table with effects of factors frequency level and voice training on produced frequencies, and maximum and minimum intensities. No significant interaction was present between the factors.
Note. df=10,539 for frequency level. df=1,548 for voice training.
* $p<0.05$

Fourier Descriptor (FD_7). The contour regularity increased significantly from 0.20 to 0.24 after voice training. The slope of the phonetogram did not change significantly.

Variable	PRE mean	PRE SD	POST mean	POST SD	t value	p
Shape						
FD_1	0.12	0.062	0.08	0.042	2.14	0.043*
FD_2	0.62	0.117	0.48	0.101	6.83	<0.001*
FD_3	0.20	0.083	0.19	0.089	0.67	0.511
FD_4	0.29	0.120	0.27	0.099	0.46	0.651
FD_5	0.17	0.086	0.15	0.075	0.85	0.404
FD_6	0.19	0.076	0.21	0.101	-0.93	0.361
FD_7	0.20	0.082	0.14	0.073	3.11	0.005*
FD_8	0.15	0.090	0.17	0.085	-0.56	0.580
FD_9	0.16	0.077	0.13	0.079	1.14	0.267
FD_{10}	0.15	0.077	0.15	0.066	-0.20	0.844
FD_{11}	0.11	0.056	0.12	0.073	-0.45	0.653
FD_{12}	0.12	0.059	0.10	0.041	0.96	0.349
Contour Regularity	0.20	0.032	0.24	0.028	-5.13	<0.001*
Slope (dB/st)	0.92	0.211	0.89	0.196	0.54	0.594
(dB/oct)	11.03	2.535	10.62	2.350	0.54	0.594
Area						
Enclosed area	0.20	0.045	0.27	0.033	-7.76	<0.001*
Freq range (# oct)	2.96	0.388	2.90	0.351	1.21	0.239
Dyn. ranges and central position						
$WDR_{mff-3\,st}$	4.39	1.783	5.04	1.389	-3.05	0.005*
WDR_{mff}	5.29	1.14	6.21	0.159	-4.25	<0.001*
$WDR_{mff+1/2\,oct}$	5.68	0.493	6.24	0.215	-5.99	<0.001*
$WDR_{mff+1\,oct}$	5.21	1.194	6.19	0.429	-5.38	<0.001*
$CP_{mff-3st}$	-0.79	0.959	-1.17	0.755	1.92	0.066
CP_{mff}	-0.10	0.798	-0.29	0.237	1.12	0.272
$CP_{mff+1/2\,oct}$	0.49	0.636	0.12	0.260	3.09	0.005*
$CP_{mff+1\,oct}$	0.99	0.862	0.52	0.539	2.87	0.008*

Table 4. Analyzed characteristics of phonetograms of a group of speech therapy students at the beginning of the study (PRE) and after two and a half years of education (POST). Mean values and standard deviations (SD) are given. FD = Fourier Descriptor, WDR = Weighted Dynamic Range, CP = Central Position, mff = mean speaking fundamental frequency, st = semitones, oct = octaves.
Note. df = 24. $^{*}p<0.05$.

Area. The enclosed area shows a highly significant increase from 0.20 to 0.27. In contrast, the frequency range did not change significantly.

WDR and CP. On all four frequency positions the WDR increases highly significantly with vocal training (see Table 4). This difference is most significant at the mean speaking fundamental frequency plus half an octave ($WDR_{mff+1/2oct}$).

Before and after vocal training, the calculated CP of the WDR show both negative and positive values, implying that the centre of the intensity range is below or above the reference intensity of 75 dB, respectively. The transition (with increasing frequency within the phonetogram) from a negative to a positive value occurs between the sampled frequencies mff and mff+1/2oct. The CP of the average trained phonetogram is located at a lower level at all four frequency intervals. The difference in location of the CP is significant at the sampled frequencies $CP_{mff+1/2oct}$ and $CP_{mff+1octave}$.

	PRE	POST	t-value	p
/s/	22.8 (10.68)	25.97 (10.86)	-1.10	0.281
/z/	20.6 (5.72)	18.2 (4.45)	2.37	0.026*
s/z ratio	1.11 (0.403)	1.45 (0.538)	-2.56	0.017*

Table 5. Mean phonation times and s/z ratio of a group of speech therapy students at the beginning of the study (PRE) and after two and a half years of education (POST). Differences between PRE and POST are expressed in t values as well as corresponding probability values (df = 24). Standard deviations are given between brackets.
* $p<0.05$

Maximum Phonation Times

Table 5 summarizes the results of produced maximum phonation times. Surprisingly, phonation time of the vocalized consonant /z/ decreased with vocal training. The average maximum phonation time of the voiceless consonant /s/, however, increased and showed the largest value. The standard deviation for the phonation time of /s/ is large in comparison with that of /z/ and expresses the difference across students in expiratory control. Paired t-tests revealed a significant decrease in phonation time for /z/ and a significant increase for the s/z ratio from 1.11 to 1.45.

DISCUSSION

Effects of voice training are discussed with respect to phonetograms, followed by a discussion of presented maximum phonation times.

Phonetograms

Clear differences were established between average phonetograms before and after vocal training. As demonstrated by the variables range, enclosed area, and Weighted Dynamic Range, voice capacities are increased with respect to dynamic ranges (see Tables 2 and 4). This increment offers enhanced possibilities to modulate intensity during speech. The lower Central Positions (see Table 4) indicate that the minimum intensities can be produced softer. The minimum intensity profile has a direct relation with threshold pressure (31, 32). Threshold pressure is, among others, determined by fundamental frequency, vocal fold thickness and longitudinal tension. A successful implementation of the Smith accent method should lead to phonation with vocal folds with increased thickness, which thus lowers threshold pressure and thereby increases soft voice capacities. Another condition favorable for producing soft phonations is increased breath support and expiratory control. This improved control should follow from the exercises of Coblenzer. A softer minimum intensity profile can therefore be regarded as a specific result of voice training.

The maximum intensity profile shows a significant increase in loudness. The sound pressure level measured outside the mouth is the product of both laryngeal source and supralaryngeal resonance characteristics. The maximum loudness at the source level is determined by the maximum subglottal pressure a subject is able -- or willing -- to produce. This maximum is limited by various factors, but most importantly is phonatory instability pressure (31). Above this pressure the voice quality deteriorates. Subjects, therefore, avoid passing this threshold pressure. Although subglottal pressure is the main component determining SPL, another important laryngeal condition promoting loud phonations is "flow phonation". Flow phonation is achieved in a relaxed mode of phonation (33). A successful application of the Smith accent method shows a positive relation between increased airflow and SPL (34). Therefore, trained subjects should be able to produce louder phonations. The trained subjects might also benefit from an increased awareness of their own voice capacities. This awareness diminishes the fear of damaging vocal structures by phonation at high sound intensities, which stimulates a deliberate exploration of loud voice capacities.

The vocal tract supplies optimal acoustical characteristics for the power transfer from voice source to mouth opening. The resonances in the vocal tract give an important contribution to the overall SPL. Various parts of the methods involved in voice training, such as the resonance method and the nasalizing method, are aimed at improving the impedance match between the glottis and free space (31).

The observed differences in dynamic range might also be attributed to the time period of two and a half years involved in this study and a possible influence of ageing of the students. Regression analysis in a data base with normative data was used to examine the effect of age on maximum and minimum intensity. In a cohort of 83 untrained female subjects (age between 17 and 25, mean age 18.9, SD = 1.61 years) a significant negative effect (p=0.0038) of age was established on maximum intensity, indicating a decrease in maximum intensity with age. A significant negative effect of age (p=0.0161) with a regression coefficient of -0.79 was found for the minimum intensity. Thus, in the investigated group of students a decrease in minimum intensity of 2.0 dB might be expected, which is much less than the measured 7.2 dB. The increase in loud and soft voice capacities can, therefore, not be explained by ageing of the students.

Pertinent studies focusing on phonetographical differences between trained and untrained subjects showed, to a varying extent, increased dynamic ranges in the trained ones. Åkerlund et al. (15) established a significantly increased loud phonation contour, while no difference was found in the soft phonation contour. In contrast, Sulter et al. (17) found increased soft phonation capacities in trained amateur singers, while there was no difference in loud phonation contour. Awan (16) established a significant increase in both loud and soft capacities. These studies also established an increase in frequency range in the trained subjects. In the present study such a difference was not found. The unchanged frequency range might be due to the fact that the training methods are not aimed at improving singing capacities, which would have implied training phonation in falsetto, a register normally not used in Dutch speech. The PRE and POST frequency ranges are in agreement with the frequency range of a large group of untrained female subjects (17).

Besides the minute analysis of the dynamic range in the speaking voice area, the phonetographical conjoint analysis method also enables a numerical evaluation of shapes of phonetograms. Studying Figure 2, the average trained phonetogram clearly shows the larger enclosed area. In both average phonetograms there is a local depression, related to a register shift, in an otherwise rising loud phonation contour; however, the location along the frequency range is shifted upward in frequency in the trained phonetogram. Three of the twelve FDs used in this study, namely FD_1, FD_2 and FD_7, show a significant difference between the PRE and POST phonetogram. The lower value for FD_1 after vocal training represents the more circular shape of the average phonetogram. As a measure for the ellipticity of a contour, the decrease in FD_2 in the trained students reflects the more rounded ends of the average phonetogram (29). Comparing untrained and trained (singing) vocal groups in another study Sulter et al. (17) established a difference in FD_3. Future studies employing the shape analysis technique should confirm that a specific change in FDs reflects the nature and content of voice training.

Maximum phonation times

Maximum phonation times can be used to obtain valuable information about the voice function and expiratory system (19). Phonation time is determined both by the lung volume employed during phonation and the regulation of the generated airstream. Although differences in lung and, thus, phonation volumes exist among subjects, as this volume depends on gender, length and age, for a given phonation volume the phonation time is mainly determined by laryngeal resistance. Because in this study the same subjects were measured before (PRE) and after (POST) voice training, biasing influences of gender and length are not present. In addition, differences in maximum phonation times of vowels and voiced consonants are based on a modified regulation of both glottal resistance and expiratory control. In voiceless consonants, the maximum time is, apart from the phonation volume, dependent on expiratory control. Regression analysis in the described control cohort revealed no significant effect of age on maximum phonation times.

The significant decrease in phonation time of the voiced consonant /z/ suggests a decrease in glottal resistance. Without the presence of vocal fold irregularities, which was confirmed laryngostroboscopically, this decrement is related to more relaxed vocal folds, one of the goals of the trained exercises and methods. Large standard deviations are found PRE and POST for the /s/, expressing the large differences in expiratory control between subjects.

Previously the s/z ratio has been used to evaluate phonatory function, and possibly detect laryngeal pathology (20), or to follow the ageing process (35). Mueller (35) also refers to the sparsity of available normative data. The values established in this study for the /s/ and /z/ of the untrained students are only slightly shorter than those published in the study of Mueller (35). The s/z ratio, established in this study for untrained students, is slightly larger than the value reported by Mueller. All values (/s/, /z/ and s/z ratio) are close to those reported in the Eckel & Boone (20) study for a combined group of normal male and female subjects. In this study, a significant increase was found in the s/z ratio from 1.11 in the untrained situation to 1.45 after voice training. This increment is based on a decrease in phonation time of /z/ with an increase in /s/ value. In their article, Eckel & Boone (20) regard a larger than "normal" s/z ratio as a sign of laryngeal pathology. This study, however, shows that s/z ratios should be used with care and always related to absolute values of /s/ and /z/.

CONCLUSIONS

Specific voice training given to a group of untrained voice therapy students resulted in changes in both voice capacities and s/z ratio.

Analysis of differences in average phonetograms showed that trained students had increased dynamic ranges in the frequency range used in normal speech, and that soft voice capacities were greatly extended. Frequency ranges remained unchanged.

The s/z ratio increased significantly to a value heretofore associated with vocal pathology. This increase is presumably caused by reduced laryngeal resistance and improved expiratory control. S/z ratios should, therefore, be given with absolute values of /s/ or /z/ to distinguish between a normal subject having received vocal training leading to voluntarily reduced laryngeal resistance and improved expiratory control, and a patient with vocal pathology resulting in involuntarily reduced laryngeal resistance.

It is hypothesized that the increased dynamic capacities in the soft voice region are caused by the reduction of the threshold pressure, which relies on specific laryngeal features, such as vocal fold thickness and length, as well as on improved neuromyogenic control over breath support. The increased loud voice capacities could result from "resonance tuning" and flow phonation.

Voice training leads to a more efficient use of inspired air and optimizes both the sound generating capacities of vocal structures, as well as the resonance capacities of the vocal tract. Persons with limited natural capacities regarding sound production can therefore benefit from these functional changes, and dysfunctional voice use leading to pathology could be prevented. Prospective studies with different groups of vocal pathology should confirm the observations made on voice healthy subjects in this study.

References

1. Brodnitz F, Fröschels E. Treatment of nodules of vocal cords by chewing method. Arch Otol 1954;59:560-5.
2. Brodnitz F. Vocal rehabilitation in benign lesions of the vocal cords. J Speech Hear Dis 1958;23:112-7.
3. Brodnitz F. Goals, Results and Limitations of Vocal Rehabilitation. Arch Otol 1963;77:148-56.
4. Fisher H, Logemann J. Objective evaluation of therapy for vocal nodules: A Case Report. J Speech Hear Dis 1966;35:277-85.
5. Lancer J, Syder D, Jones A, Le Boutillier A. Vocal cord nodules: a review. Clin Otol 1988;13:43-51.
6. Murry T, Woodson G. A comparison of three methods for the management of vocal fold nodules. J Voice 1992;6:271-6.
7. Damsté P. Oefeningen voor volledige sluiting van de glottis als onderdeel van stemtherapie. Logop Foniatr 1973;45:205-9.
8. Yamaguchi H, Yotsukura Y, Sata H, Watanabe Y, Hirose H, Kobayashi N, Bless D. Pushing exercise program to correct glottal incompetence. J Voice 1993;7:250-6.
9. Stemple J, Lee L, D'Amico B, Pickup B. Efficacy of vocal function exercises as a method of improving voice production. J Voice 1994;8:271-8.
10. Brodnitz F. Post-Operative vocal rehabilitation in benign lesions of the vocal cords. Folia Phoniatr 1955;7:193-200.

11. Brodnitz F. Contact Ulcer of the Larynx. *Arch Otol* 1961;74:70-80.
12. Lancer J, Syder D, Jones A, Le Boutillier A. The outcome of different management patterns for vocal cord nodules. *J Lar Otol* 1988;102:423-7.
13. Murry T, Caligiuri M. Phonatory and nonphonatory motor control in singers. *J Voice* 1989;3:257-63.
14. Schutte H. Over het fonetogram. *Logop Foniatr* 1975;47:82-92.
15. Åkerlund L, Gramming P, Sundberg J. Phonetogram and averages of Sound Pressure Levels and fundamental frequencies of speech: Comparison between female singers and nonsingers. *J Voice* 1992;6:55-63.
16. Awan S. Phonetographic profiles and F0-SPL characteristics of untrained versus trained vocal groups. *J Voice* 1991;5:41-50.
17. Sulter A, Schutte H, Miller D. Differences in phonetogram features between male and female subjects with and without vocal training. *J Voice* 1995;9:363-77.
18. Awan S. Superimposition of speaking voice characteristics and phonetograms in untrained and trained vocal groups. *J Voice* 1993;7:30-7.
19. Kent R, Kent J, Rosenbek J. Maximum performance tests of speech production. *J Speech Hear Dis* 1987;52:367-87.
20. Eckel F, Boone D. The s/z ratio as an indicator of laryngeal pathology. *J Speech Hear Dis* 1981;46:147-9.
21. Seidner W, Wendler J. *Die Sängerstimme*. Berlin: Henschelverlag Kunst und Gesellschaft, 1982.
22. Coblenzer H, Muhar F. Die Phonationsatmung. *Wiener Klinische Wochenschrift* 1965;77:946-53.
23. Coblenzer H, Muhar F. *Atem und Stimme*. Wien: Paul Gerin, 1976.
24. Veldkamp K. *De techniek van het spreken*. Groningen: Wolters-Noordhoff, 1965.
25. Eldar A. *Spreken en zingen*. Assen: Van Gorcum, 1976.
26. Smith S. *Die Akzentmethode und ihre theoretische Voraussetzungen*. Flensburg: Spezial-Pädagogische Verlag KG, 1980.
27. Pahn J, Pahn E. Die Nasalierungsmethode. *Logop Foniatr* 1986;58:126-32.
28. Schutte H, Seidner W. Recommendation by the Union of European Phoniatricians (UEP): Standardizing Voice Area Measurement/Phonetography. *Folia Phoniatr* 1983;35:286-8.
29. Sulter A, Wit H, Schutte H, Miller D. A structured approach to voice range profile (phonetogram) analysis. *J Speech Hear Res* 1994;37:1076-85.
30. Zahn C, Roskies R. Fourier Descriptors for plane closed curves. *IEEE Trans Comp* 1972;21:269-81.
31. Titze I. Acoustic interpretation of the voice range profile (phonetogram). *J Speech Hear Res* 1992;35:21-34.
32. Titze I. Phonation threshold pressure: A missing link in glottal aerodynamics. *J Acoust Soc Am* 1992;91:2926-35.
33. Sundberg J, Titze I, Scherer R. Phonatory control in male singing: a study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source. *J Voice* 1993;7:15-29.
34. Kotby M, Shiromoto O, Hirano M. The Accent method of voice therapy: effect of accentuations on F0, SPL, and airflow. *J Voice* 1993;7:319-25.
35. Mueller P. Maximum durations of /s/ and /z/ and the s/z ratio of elderly subjects. *Voice Forum* 1993;2:49-54.

10

Summary and Conclusions

The basis of voice production continues to be an important topic of investigation. During phonation pressure differences are generated by vibrating vocal folds, and this process is influenced by many physiological and anatomical variables. However, details of voice production are yet to be disclosed: normative data regarding voice production are scarce, which impedes qualification of vocal characteristics. Also, limited information is available about the relation between aspects of voice production, such as voice physiology and perception of voice. Theoretical approaches of voice production with analytical models are hampered by lacking information on both average voice characteristics in a population, as well as on the constraints of certain characteristics. This study was carried out to produce normative data on voice production, as well as to qualify vocal characteristics regarding gender and status of vocal training.

Chapter 2

The larynx with vibrating vocal folds as the generator of vocal sound can be closely observed with videolaryngostroboscopy. Aspects of laryngeal appearance (larynx/pharynx ratio; epiglottal shape; asymmetry of the arytenoid region; compensatory adjustments with increasing intensity; thickness, width, length and elasticity of vocal folds) and glottal functioning (amplitudes of vocal fold excursion; duration, percentage and type of vocal fold closure; phase differences; location of glottal chink) of 214 subjects were evaluated by three judges using a standardized scaling instrument. Regarding laryngeal appearance, men were rated having a significantly higher larynx/pharynx ratio and showing more compensatory adjustments while changing sound intensity level, as compared to women. A more deviant-shaped epiglottis was observed in men, while the presence of an omega- shaped epiglottis was an exclusive finding in the male group. Compared to those of females, male vocal folds were rated thicker in the vertical dimension, smaller in the lateral dimension, longer and more tense, with smaller amplitudes of excursion during vibration. Glottal closure in male subjects was rated more complete, but briefer in duration. Subjects having received vocal training showed more often a complete glottal closure compared to untrained subjects, and were rated as having more frequently lateral phase differences of vocal fold excursion. Important influences of the variables pitch, sound intensity and age of the subject on vocal fold appearance and glottal functioning were ascertained. The results demonstrate that in clinical practice during videolaryngostroboscopical evaluation the investigator should be aware of the observed effects of presented variables on laryngeal characteristics for a proper qualification of the images.

Chapter 3

Incomplete glottal closure during phonation is associated with perceived breathiness. Breathy voices are restricted regarding voice capacities and are less resistant to voice demanding tasks. Therefore robustness of the voice source is related with glottal closure. To quantify glottal closure a frame of reference was created by investigating 47 healthy men and 92 healthy women without vocal complaints using videolaryngostroboscopy. Observing recorded images the degree of glottal closure was rated with a percentage. Results indicate that men have better glottal closure than women. An increase in vocal intensity is related to improved glottal closure, and in women a negative relationship was established between pitch and glottal closure. Normal glottal closure in men is a complete closure, whereas in women a closure of at least 90% should be attained. If these percentages cannot be established during loud phonation, it suggests the presence of a less robust larynx. To evaluate and quantify the function of the voice source, in clinical practice the larynx should not be observed at only one intensity level, but at a variety of intensity and frequency levels.

Chapter 4

During phonation vibrating vocal folds cause variations in the expiratory airflow. These variations can be measured as glottal volume velocity waveforms and they are the basis of the audible signal on which speech depends. Glottal volume velocity waveform characteristics of 224 subjects, categorized in four groups according to gender and vocal training, were determined, and their relations to sound intensity, fundamental frequency, intra oral pressure and age were analyzed. Waveforms were characterized with flow-based and time-based parameters, as well as with derived parameters. Few statistically significant differences in parameters were found between trained and untrained subjects: maximum flow declination rate is higher and the time period in which vocal folds close within a glottal cycle is shorter in trained women, compared to the untrained ones, while the waveform shows a more asymmetrical shape in trained men. Several significant differences were found between men and women. Higher flow-based parameter values were established in men. The time period in which maximum flow reduces to minimum flow in each glottal cycle was measured to be shorter in men, while the time period that vocal folds are assumed to be closed was longer, as compared to parameter values of women. These differences reflect the anatomical differences between male and female larynges. Variables such as sound intensity, intra oral pressure, pitch and age showed to have important influences on glottal waveform characteristics. Chapter 4 gives a detailed description of these influences. This description indicates that variation of each of these variables is closely related to significant changes in voice physiology, and stresses the dynamic events taking place at the glottal level during speech.

Chapter 5

Closure of the vocal folds causes abrupt cessation of the glottal flow and results in audible pressure changes. The sound intensity level is related to the abruptness of closure. This abruptness can be measured as the maximum flow declination rate of the glottal volume velocity waveform. However, the abruptness of closure is also associated with vocal fold damage. In this chapter the specific relation between sound intensity and maximum flow declination rate is mathematically described in particular groups with differing resistance to vocal fatigue. In the groups with vocal problems the abruptness of closure was higher compared to groups without problems. The given mathematical description might be used to predict susceptibility to vocal problems.

Chapter 6

Pressure changes generated in the larynx are modified in the vocal tract. The vocal tract contains the space between the glottis and the lips, and functions as a frequency selective amplifier and attenuator. By articulatory movements a meaningful dynamic modification of the voice source signal is produced and the result is speech. Speech characteristics of 224 subjects were perceptually evaluated with a bipolar semantic scaling instrument to study differences between men and women, as well as between subjects with and without vocal training. Compared to male speech, speech of women was rated more expressive and melodious, higher and clearer on the one hand, and more unpleasant, softer, shriller and weaker on the other. The ratings reflect the higher appreciation of the intonation pattern of female speech, that was further perceived as less dynamic and of a lower quality. Speech of trained subjects was more highly appreciated regarding articulation in comparison with untrained subjects, and it was also perceived as more clear. Trained subjects are therefore assumed to exploit more fully their control over the voice source and articulatory organs during speech.

Chapter 7

Phonetograms offer a clear representation of the individual voice capacities in the frequency and sound intensity range. Because of its two-dimensional character, difficulties arise when making comparisons between phonetograms. Former solutions of this problem were based on standardizing phonetograms for frequency range, thereby distorting the unique shape of each individual phonetogram. A new method to analyze phonetograms is introduced in this chapter. The structured analysis is based on quantitatively determining the features: shape, area and "speaking range" dynamics, without distorting the shape of phonetograms. The parameter sets describing these features are calculated independently of fundamental frequency. Apart from making it possible to compare phonetograms, this method also provides a tool for establishing normative data for specified groups.

Chapter 8

Voice capacities of groups, created according to gender and status of vocal training, are compared, using two different methods. One is based on the rescaling of phonetograms, while the other derives analytic variables from the features shape, area and dynamic range. Regarding gender, analysis showed that male subjects were able to produce softer phonations, whereas female subjects produced louder phonations at specific parts of their comparable frequency ranges. Trained subjects had a larger enclosed area of the phonetogram, which was primarily based on extended soft voice capabilities in both genders, and the significantly larger frequency range in trained female subjects. The shape analysis, performed with Fourier Descriptors, revealed differences for the factors gender and training.

Chapter 9

In former chapters comparisons were made between groups with and without vocal training. In this chapter the results of a prospective study, analyzing the effect of two and a half years of voice training on both phonetograms and maximum phonation times have been presented. Phonetogram analysis showed a significant improvement in both loud and soft intensity range, whereas no change in frequency range was observed. As an indicator of laryngeal functioning, the s/z ratio, which is determined by taking the ratio of the phonation times of the voiceless consonant /s/ and the voiced consonant /z/, increased significantly after vocal training, presumably reflecting a reduced laryngeal resistance and an improved expiratory control.

With information on the variation of voice quality features in a large population, data bases are constructed, which can be used as reference material in clinical practice, as well as for both models and future studies.
Large differences in laryngeal characteristics were found between men and women, reflecting the gender specific anatomy of the larynx and the associated physiology of phonation. This implies that a separate evaluation of female and male voices should be performed, and that female laryngeal and glottal characteristics are not simply comparable to male characteristics transposed by one octave.
Minor differences in laryngeal and glottal characteristics were established between untrained and trained groups. The most important difference was the more complete glottal closure in trained subjects. The minor differences imply that the vocal apparatus basically does not differ between the two groups, which might be explained by the relatively low level of training in the "trained" group. However, large differences were found between the untrained and trained groups regarding phonetogram features and perceptive ratings of speech. These differences demonstrate the better control over the voice source and articulatory organs in trained subjects. Trained subjects are, therefore, expected to exploit more fully their phonatory capacities compared to untrained subjects.

11

Samenvatting en Conclusies

Tijdens stemgeving worden drukwisselingen gegenereerd door trillende stemplooien en dit proces wordt beïnvloed door vele fysiologische en anatomische variabelen. De basisprincipes van stemgeving zijn onderwerp geweest van talrijke studies. Essentiële details betreffende stemgeving dienen echter nog te worden ingevuld: normatieve gegevens over stemvorming zijn schaars, waardoor kwalificatie van stemkarakteristieken wordt belemmerd. Daarnaast zijn weinig gegevens bekend over de relatie tussen fysiologische processen en cognitieve aspecten van het stemgeluid. De theoretische benadering van stemvorming middels analytische modellen wordt bemoeilijkt door onvoldoende gegevens over zowel “gemiddelde” stemkarakteristieken in een populatie, als ook over de grenswaarden van deze karakteristieken. De onderhavige studie werd dan ook uitgevoerd om normatieve gegevens over stemgeving vast te stellen en om stemkarakteristieken te kwalificeren op grond van geslacht en geoefendheid van de stemgebruiker.

Hoofdstuk 2

Het strottehoofd met trillende stemplooien als generator van stemgeluid kan nauwkeurig worden bestudeerd met behulp van videolaryngostroboscopie. Aspecten van de uiterlijke verschijning van het strottehoofd (larynx/pharynx ratio; vorm van de epiglottis; asymmetrie in het arytenoïdgebied; compensatoire activiteit bij toenemende geluidsintensiteit; dikte, breedte, lengte en elasticiteit van de stemplooien) en van de glottisfunctie (amplitude van de stemplooibeweging; duur, percentage en type betreffende stemplooisluiting; faseverschillen; locatie van een glottaal lek) van 214 proefpersonen werden geëvalueerd door drie juryleden met behulp van gestandaardiseerde schalen. Wat betreft de verschijningsvorm werden bij mannen een hogere larynx/pharynx ratio alsmede frequenter optredende compensatoire activiteiten bij intensiteitsveranderingen waargenomen in vergelijking met vrouwen. Een afwijkende vorm van de epiglottis werd vaker gezien in mannen, terwijl de infantiele epiglottis exclusief in mannen kon worden vastgesteld. In vergelijking met de stemplooien van vrouwen, werden de mannelijke stemplooien als dikker, smaller, langer en meer gespannen beschreven, met kleinere uitslagen van beweging. Stemplooisluiting werd als meer compleet beschreven in mannen, echter tegelijkertijd korter van duur. In getrainde proefpersonen werd vaker dan in ongetrainden een volledige stemplooisluiting waargenomen hetgeen ook opging voor faseverschillen in zijwaartse stemplooibeweging. Daarnaast werd de belangrijke invloed van de variabelen toonhoogte, geluidsintensiteit en leeftijd van de proefpersoon vastgesteld op zowel verschijningsvorm van de stemplooien als glottisfunctie. De resultaten geven aan dat de onderzoeker zich tijdens videolaryngostroboscopische evaluatie in de klinische praktijk bewust moet zijn van de invloeden van de gepresenteerde variabelen om een adequate beoordeling van de stem te kunnen uitvoeren.

Hoofdstuk 3

Stemmen met een onvolledige stemplooisluiting worden gekenmerkt door een zekere mate van heesheid. Hese stemmen zijn beperkt qua stemmogelijkheden en minder bestand tegen intensieve stemopdrachten. Belastbaarheid van de stem is dan ook gerelateerd aan stemplooisluiting. Om stemplooisluiting te kunnen kwalificeren werd een referentiekader gevormd door 47 gezonde mannen en 92 gezonde vrouwen zonder stemklachten te onderzoeken middels videolaryngostroboscopie. De opgenomen beelden werden bestudeerd en de sluiting in percentages uitgedrukt. De resultaten tonen dat de sluiting bij mannen beter is dan bij vrouwen. Toename van de geluidsintensiteit gaat samen met verbetering van de sluiting, en bij de vrouwen werd een negatief verband gelegd tussen toonhoogte en stemplooisluiting. Een normale stemplooisluiting is in mannen een complete sluiting, terwijl bij vrouwen een sluiting van tenminste 90% moet worden vastgesteld. Als dergelijke sluitingen niet worden bereikt bij luide fonatie dan is er sprake van een minder belastbaar strottehoofd. Om de functionaliteit van de stembron te evalueren en te kwantificeren dient het strottehoofd in de klinische praktijk niet slechts op één intensiteitsniveau te worden geobserveerd, maar tijdens meerdere intensiteiten en toonhoogten.

Hoofdstuk 4

Tijdens stemgeving veroorzaken trillende stemplooien wisselingen in de uitademingslucht. Deze wisselingen kunnen worden gemeten als glottale luchtstroomgolfvormen en deze zijn de basis van het hoorbare signaal waarop spraak berust. Golfvormkarakteristieken werden bepaald van 224 proefpersonen, welke waren ingedeeld in vier groepen naar geslacht en geoefendheid, en relaties met geluidsintensiteit, toonhoogte, intra-orale druk en leeftijd werden geanalyseerd. Golfvormen werden gekarakteriseerd door middel van luchtstroom- en temporele parameters, alsmede hiervan afgeleide parameters. Slechts enkele statistisch significante verschillen werden vastgesteld tussen ongetrainde en getrainde proefpersonen: de maximale afname van de luchtstroomsnelheid is groter, en de sluitingsduur van de stemplooien korter in getrainde vrouwen, terwijl de golfvorm een meer asymmetrisch aspect heeft in de getrainde mannen. Meerdere significante verschillen werden gevonden tussen mannen en vrouwen. Hogere waarden voor de luchtstroomparameters werden gemeten in mannen. Bij mannen is de tijd waarin de maximale luchtstroom afneemt tot minimale luchtstroom korter en de tijdsduur waarin de stemplooien gesloten zijn langer, in vergelijking met vrouwelijke parameterwaarden. Deze verschillen weerspiegelen de anatomische verschillen tussen het mannelijke en vrouwelijke strottehoofd. Variabelen zoals geluidsintensiteit, intra-orale druk, toonhoogte en leeftijd hebben een belangrijke invloed op golfvormkarakteristieken. Hoofdstuk 4 geeft een gedetailleerde beschrijving van deze invloeden. Hieruit blijkt dat variatie van ieder van de variabelen samenhangt met betekenisvolle veranderingen in stemfysiologie en benadrukt daarmee de dynamische gebeurtenissen die zich voltrekken op stemplooiniveau tijdens spraak.

Hoofdstuk 5

Sluiting van de stemplooien veroorzaakt een plotselinge vermindering van de glottale luchtstroom, hetgeen hoorbare drukwisselingen oplevert. De geluidsintensiteit is gerelateerd aan de abruptheid waarmee de sluiting zich voltrekt. De abruptheid kan worden gemeten als de maximale afname van de luchtstroomsnelheid (MFDR) van de glottale luchtstroomgolfvorm. De abruptheid is echter ook geassocieerd met beschadiging van de stemplooien. In dit hoofdstuk is de specifieke relatie tussen geluidsintensiteit en MFDR mathematisch beschreven in groepen met verschillende tolerantie tot vermoeibaarheid van de stem. In de groep met stemproblemen werden meer abrupte sluitingen gevonden dan in de groepen zonder stemproblemen. De toegepaste mathematische beschrijving zou gebruikt kunnen worden om vatbaarheid voor stemproblemen te voorspellen.

Hoofdstuk 6

De in het strottehoofd gegenereerde drukwisselingen worden gemodificeerd in het aanzetstuk. Het aanzetstuk omvat de ruimte tussen stemplooien en lippen, en werkt als een frequentie-selectief filter. Door articulatorische bewegingen wordt het basissignaal betekenisvol gemodificeerd en het resultaat is spraak. Spraakkenmerken van 224 proefpersonen werden perceptief geëvalueerd met behulp van bipolaire semantische schalen, om daarmee zowel verschillen tussen mannen en vrouwen, als tussen proefpersonen met en zonder stemtraining te bestuderen. Vergeleken met spraak van mannen werd spraak van vrouwen aan de ene kant meer expressief en melodieus, hoger en helderder beoordeeld, maar aan de andere kant onplezieriger, zachter, scheller en zwakker. De beoordelingen weerspiegelen de hogere waardering voor vrouwelijke intonatiepatronen, echter tegelijkertijd geven zij ook de kleinere dynamiek aan van vrouwelijk spraak die voorts gekenmerkt is door een lagere kwaliteit. In vergelijking met spraak van ongetrainden werd spraak van getrainde proefpersonen beter gewaardeerd wat betreft articulatie, en deze spraak kwam ook helderder over. Getrainde proefpersonen lijken daarmee hun mogelijkheden wat spraak betreft beter te benutten door een betere controle over zowel de stembron als de articulatorische organen.

Hoofdstuk 7

Fonetogrammen bieden een goede weergave van individuele stemmogelijkheden qua melodisch en luidheidsbereik. Vanwege het twee-dimensionale karakter is het moeilijk om vergelijkingen te maken tussen fonetogrammen. Oplossingen hiervoor werden gezocht in standaardisatie van de fonetogrammen voor het melodisch bereik, waarbij de unieke vorm van ieder individueel fonetogram vervolgens op niet eenduidige wijze wordt aangepast. In dit hoofdstuk is een nieuwe methode om fonetogrammen te analyseren geïntroduceerd. De gestructureerde analyse is gebaseerd op een kwantitatieve bepaling van de eigenschappen: vorm, oppervlakte en dynamische spraakomvang, zonder dat daarbij de vorm van de fonetogrammen wordt veranderd. De parameterverzamelingen die de eigenschappen beschrijven worden onafhankelijk van de grondtoon berekend. Naast de mogelijkheid om de methode te gebruiken om fonetogrammen te vergelijken kan deze ook worden gebruikt om normatieve gegevens voor specifieke groepen te produceren.

Hoofdstuk 8

In dit hoofdstuk zijn stemmogelijkheden van groepen, ingedeeld naar geslacht en eventuele stemtraining, met elkaar vergeleken met behulp van twee verschillende methoden. De eerste is gebaseerd op standaardisatie van de frequentieschaal van fonetogrammen, terwijl de andere berust op verwerking van variabelen welke resulteren na analyse van de eigenschappen vorm, oppervlakte en dynamische spraakomvang. Mannen bleken in staat te zijn zachter geluid te maken, terwijl de vrouwelijke proefpersonen op specifieke frequentiegebieden van het fonetogram luider konden foneren. Getrainde poefpersonen hadden een fonetogram met een grotere oppervlakte, hetgeen te maken had met toegenomen zachte stemmogelijkheden in beide geslachten en het significant toegenomen melodisch bereik in de getrainde vrouwelijke proefpersonen. De analyse van de vorm, uitgevoerd met Fourier Descriptoren, gaf verschillen aan, zowel tussen de geslachten als tussen ongetrainden en getrainden.

Hoofdstuk 9

In de vorige hoofdstukken werden vergelijkingen gemaakt tussen groepen met en zonder stemtraining. In dit hoofdstuk zijn de resultaten gegeven van een prospectieve studie, waarin de invloed van twee en een half jaar stemtraining op het fonetogram en de fonatietijden is geanalyseerd. Analyse van de fonetogrammen liet zowel een toename van de zachte als de luide stemmogelijkheden zien, terwijl geen verandering van melodisch bereik werd geconstateerd. Stemfunctie kan worden geëvalueerd met de s/z ratio. Deze wordt bepaald door de ratio te nemen van de fonatietijden van de stemloze medeklinker /s/ en de stemhebbende medeklinker /z/. De s/z ratio nam significant toe na de stemtraining, hetgeen waarschijnlijk berust op een afname van de weerstand van de stemplooien tijdens stemgeving in combinatie met een toegenomen controle over de uitademing.

Met informatie over de variatie van stemkwaliteitskenmerken in een grote populatie kunnen databases worden gemaakt, welke toegepast kunnen worden ten behoeve van modelvorming en toetsing, alsmede ter ontwikkeling van normatieve gegevens voor toekomstige studies. Tevens kunnen dergelijke gegevens gebruikt worden als referentiekader in de klinische praktijk.
Significante verschillen in strottehoofdeigenschappen werden vastgesteld tussen mannen en vrouwen. Deze verschillen weerspiegelen de geslachtsspecifieke anatomie van het strottehoofd en de hieraan gerelateerde fysiologie van de stemgeving. Na frequentiecorrectie zijn de strottehoofd- en stemplooi-eigenschappen van de vrouw niet vergelijkbaar met die van de man. Daarom dient evaluatie van de stem onafhankelijk voor mannen en vrouwen te geschieden.
Geringere verschillen in strottehoofd- en stemplooi-eigenschappen werden gevonden tussen de ongetrainde en getrainde groepen. Het belangrijkste verschil was de completere stemplooisluiting in de getrainde proefpersonen. De geringe verschillen wijzen erop, dat het stemapparaat in essentie niet verschilt tussen de groepen. Echter grote verschillen tussen de ongetrainde en getrainde groep werden wel gevonden in fonetogrameigenschappen en in de perceptieve

beoordeling van spraak. Deze verschillen wijzen op de betere controle over stem- en spraakorganen in de getrainde proefpersonen, waardoor de aanwezige mogelijkheden om geluid te produceren optimaal kunnen worden benut.

Appendix: Laryngostroboscopic rating form - Explanation of scales

Laryngeal appearance

1 Larynx/pharynx ratio (three-point scale): 1 = large, 2 = normal, 3 = small.
2 Epiglottal shape (five-point scale): 1 = large, 2 = normal, 3 = small, 4 = omega shaped, 5 = deviant (other) shaped.
3 Asymmetry arytenoid region (four-point scale): 1 = severe asymmetry, 2 = asymmetry, 3 = slight asymmetry, 4 = no asymmetry.

Compensatory adjustments (four-point scale): 1 = clearly visible, 2 = visible, 3 = almost absent, 4 = not visible.

Vocal fold appearance (five-point scales)

1 Thickness: 1 = thin, 5 = thick
2 Width: 1 = small, 5 = broad
3 Length: 1 = short, 5 = long
4 Elasticity: 1 = slack, 5 = tense

Glottal functioning

1 Amplitudes (four-point scale): 1 = large, 2 = normal, 3 = small, 4 = not visible
2 Vertical, horizontal and lateral phase differences (two-point scales): 0 = not visible, 1 = visible
3 Duration of closure (four-point scale): 1 = normal, 2 = shortened, 3 = short, 4 = no closure
4 Closure type (after Södersten, ref 29 Chapter 2), two main categories (see Figure 1, Chapter 2).

Category I (six-point scale): 1 = complete closure, 2 = incomplete closure in the cartilaginous part, 3 = triangular incomplete closure anterior to the vocal processes, 4 = triangular incomplete closure of the posterior thirds of the folds, 5 = incomplete closure of the posterior two thirds of the folds, 6 = incomplete closure all along the folds.

Category II (four-point scale): A = spindle-shaped incomplete closure, closure at the vocal processes, B = spindle-shaped incomplete closure at the posterior thirds of the folds, closure at the vocal processes, C = spindle-shaped incomplete closure at the anterior third of the folds, closure at the vocal processes, D = spindle-shaped incomplete closure at the posterior and the anterior thirds of the folds, closure at the vocal processes and at the middle of the membranous portion.

5 Percentage of closure (a figure has to be given): an estimation of the relative length of the closed portion of the glottis in the most closed phase,

0% = no closure, 100% = complete closure.

6 Location of chink(three-point scale): 1 = membranous, 2 = cartilaginous, 3 = not visible.

Dankwoord

Na jaren hard weken is het beoogde resultaat dan toch bereikt: het proefschrift is daar! Velen hebben een bijdrage geleverd in de totstandkoming van het “boekje”. Op het gevaar af enkelen te vergeten, wil ik een aantal mensen met name bedanken voor hun inspanningen.

Prof. dr. H.K. Schutte, beste Harm. Je wist me in een eerste bezoek enthousiast te krijgen voor stemonderzoek en ik ben dat enthousiasme nimmer kwijt geraakt. Ziehier het resultaat van jaren onderzoekswerk en van onze gezamenlijke inspanningen. Met veel waardering en dankbaarheid denk ik terug aan onze werkzaamheden. Heel veel dank voor alles wat je me hebt bijgebracht.

D.G. Miller, beste Don. De eerste schreden van het wetenschappelijke schrijverspad konden mede door jouw routebeschrijvingen worden gezet. Daarnaast nam je ook nog plaats in de bezemwagen en zorgde na afdoende mentale ravitaillering voor een frisse nieuwe start. Dit proefschrift zou zonder jouw steun niet zijn verschenen. Ik ben je dan ook veel dank verschuldigd.

Prof. dr. F.W.J. Albers. Gegeven de omstandigheden ben ik u dankbaar voor het feit dat u als promotor wilde fungeren en de afronding van het proefschrift wilde begeleiden.

Prof. dr. ir. H.P. Wit, beste Hero. Verzot als je bent op emissies zorgde ik regelmatig voor nieuwe varianten in de vorm van theoretische vraagstukken. Jouw antwoorden en praktische oplossingen gaven me nieuwe mogelijkheden door te gaan met het onderzoek. Dank voor al je inspanningen.

Prof. dr. P.H. Dejonckere, prof. dr. J.L.N. Roodenburg, prof. dr. F. Zwarts. Dank voor het feit dat u zitting heeft genomen in de promotiecommissie.

Prof. dr. J.M. Minderhoud. Door uw inzet jaren geleden kon ik als uitvoerder van het onderzoeksproject starten. Nogmaals mijn dank daarvoor.

Dr. H.F.M. Peters, beste Herman. Als medebegeleider hebben we menigmaal constructief overleg gehad, waarvoor dank.

Prof. dr. L. Boves en dr. ir. B. Cranen, beste Loe en Bert. Op fundamenteel wetenschappelijk gebied kon ik te allen tijde met jullie opbouwend van gedachte wisselen, daarbij fungeerden jullie als direct oproepbare “achterwachten”. Ik heb de samenwerking in de afgelopen jaren zeer gewaardeerd.

Alle ruim tweehonderd proefpersonen persoonlijk bedanken voert te ver. Hierbij wil ik allen heel hartelijk danken voor hun deelname.

Ch. Verhoog, E. Micka, L. Bors, R. De Vries en J. Dijkstra worden met name bedankt voor het leggen van de contacten met de proefpersonen.

Dr. S.M. Goorhuis-Brouwer en M. Reinen Peters wil ik bedanken voor de geboden mogelijkheid een jury voor het perceptieve experiment te formeren.

Prof. J.S. Perkell, Ph.D., E.B. Holmberg, Ph.D., R.E. Hillman, Ph.D., prof. K. Stevens, Ph.D. and J. Wozniak. My stay at the Speech Group, R.L.E., M.I.T. has had a significant influence on my scientific education. Discussions with members of your group gave inspiration for a further progress of the research project. I would like to thank you all for your hospitality and for offering me the possibility to perform an experiment on vocal modes at your lab.

B. and H. Heesch, dear Barbara and Harry. Your hospitality made my stay at Boston unforgettable, our friendship will last forever.

Drs. G.L.J. Nieboer, beste George. Ik heb dankbaar gebruik kunnen maken van jouw kennis bij het opzetten van enkele van de diverse onderzoeksprotocollen.

Dr. R. van den Brink en drs. R. Meijer, beste Rob en Rinze. Dank voor jullie statistische adviezen. Uit de diverse hoofdstukken blijkt wel hoe belangrijk de statistiek voor de verwerking en presentatie van de resultaten is geweest.

J. Meijer, beste Jorien. Jouw afstudeeronderzoek heeft de basis gelegd voor een van de hoofdstukken. Met veel plezier denk ik terug aan onze samenwerking.

Drs. G. de Jong en drs. A. Reichenbach, beste Gea en Anthonie. Door jullie inzet en ijverige computerwerk kreeg ik een instrument aangereikt, waar ik alle data op plezierige wijze in kwijt kon. Hiermee werd het fundament gelegd voor verdere verwerking van de gegevens. Het enthousiasme waarmee het een en ander voor elkaar werd gekregen heb ik zeer gewaardeerd.

Drs. G.A. Sulter, beste Geert. Je bereidheid om op wisselende, ongebruikelijke en onregelmatige tijden te assisteren bij de vergaring en invoering van de gegevens vond ik geweldig. Ik had bij de aanstelling geen betere keuze kunnen maken. Ik hoop dat we in de toekomst nog eens een vervolg kunnen geven aan onze gezamenlijke onderzoeksinspanningen.

Ing. J.H. van Dijk en F. Leemhuis, beste Jan en Frits. Dank voor de geboden technische bijstand.

Ing. F. de Haan en G. Hasselman, beste Frank en Gijs. Mede door jullie programmeerwerk konden fysiologische signalen en fonetogrammen nader geanalyseerd worden.

Lieve Willy. Je wilde niet voor in het proefschrift worden genoemd. Dat weerspiegelt je bescheidenheid, want indirect heb je ook een belangrijk aandeel gehad in de totstandkoming van het proefschrift. De laatste maanden moest zelfs de “heilige” zaterdag eraan geloven. Ik hoop met de afronding van het proefschrift meer tijd te hebben voor onze vruchtbare relatie.

Lieve Heit en Mem. Met recht ben ik trots op jullie. De mogelijkheid te kunnen studeren, alsmede de niet aflatende belangstelling en stimulering hebben ervoor gezorgd, dat wederom een langdurige studie met een promotie kan worden afgesloten

Last, but not least de paranimfen dr. R.F.E. Wolf en drs. S.J. Ham, beste Rienhart en John. Jaren geleden kondigden jullie reeds aan in de markt te zijn als “paraneef” voor begeleiding van jullie oom in het Academiegebouw. Ik ben jullie zeer dankbaar dat jullie de daad bij het woord hebt willen voegen.