The Pitch Rise Paradigm: A New Task for Real-Time Endoscopy of Non-Stationary Phonation

Oliver Rasp\textsuperscript{a}  Jörg Lohscheller\textsuperscript{a}  Michael Doellinger\textsuperscript{c}  Ulrich Eysholdt\textsuperscript{a}  Ulrich Hoppe\textsuperscript{b}

Departments of \textsuperscript{a}Phoniatrics and \textsuperscript{b}Audiology, University of Erlangen-Nuremberg, Erlangen, Germany; \textsuperscript{c}Laryngeal Dynamics Laboratory, Division of Head and Neck Surgery, UCLA School of Medicine, Los Angeles, Calif., USA

Key Words
High-speed camera  •  Pitch rise  •  Vocal fold oscillations  •  Glottal closure  •  Quantitative laryngoscopy

Abstract
As standard stroboscopy is restricted to the recording of periodic vocal fold vibrations, observations of non-stationary laryngeal mechanisms demand real-time recording systems, the most advanced being the high-speed video technique. It allows the registration of laryngeal parameters during a variation of the fundamental frequency. The aim of this study was to compare amplitude and frequency parameters of vocal fold vibration during stationary and non-stationary phonation, i.e. a monotonous pitch rise. Twenty-nine young female adults with no incidence of voice disorders were examined while performing two different phonation tasks: sustained phonation with a constant frequency and a monotonous pitch rise. Endoscopic recordings and the acoustic signals were acquired simultaneously. Both acoustic and laryngeal parameters were derived for short time intervals of 17.8 ms for the constant pitch and pitch rise conditions. Instantaneous frequency, sound pressure level, vibratory amplitudes of the vocal folds and the type of glottal closure were compared. At the beginning of the pitch rise, the acoustic and laryngeal parameters were similar to the parameters that occurred within the sustained phonation conditions. In contrast, the laryngeal parameters at the middle and at the end of the pitch rise differed substantially from those during sustained phonation. For the first time, quantitative measures of the growing glottal chink and the vibration amplitude decrease during pitch increase could be taken. In general, the image evaluation of the pitch rise paradigm can be subdivided into the starting, the raising and the...
As each phase can be considered as quasi-stationary, existing software modules are capable of analysing the process by treating each phase separately. Hence, the pitch rise condition may be suitable for clinical examination to detect information of voice disturbances that cannot be visualized during sustained phonation.

Introduction

Acoustic communication is based on the production of speech signals. Technically, voice constitutes the carrier signal of speech, to which information is imposed by modulation, i.e. articulation. Running speech is characterized by a number of transient processes and by instantaneous synchronous adjustments of articulatory and glottal parameters. A reasonable part of information transmission is encoded in prosody or intonation, which are, among others, important parameters both for speech intelligibility and transfer of context and emotional aspects. Prosody is primarily controlled by variation of voice pitch, which corresponds to a change in vibration frequency of the vocal folds. Pitch is determined mainly by length, tension and mass of the vibrating vocal folds [1], while subglottal pressure has only a marginal influence [1, 2]. Since the control of vibrating mass, vocal fold length and vocal fold tension cannot be separated, the details of the underlying mechanisms for pitch variation are still unknown. Therefore, the investigation of laryngeal mechanisms corresponding to isolated pitch variations (i.e. pitch rise or decline) may provide important information about voice and speech performance in everyday life.

The conventional visualization technique of vocal fold vibrations is stroboscopy, which allows the investigation of periodic vocal fold oscillations occurring during sustained phonation [3, 4]. However, due to violation of the Shannon sampling theorem, stroboscopy is not suitable for the investigation of aperiodic transient vocal fold movements including irregularities or short-time pitch variations [5, 6]. Observation of aperiodic movements needs real-time recording techniques, the most advanced technique being digital high-speed glottography [5, 7]. Current systems are capable of recording up to 10,000 video frames/s. In particular, they are suited to analyse non-stationary effects occurring in a large variety of physiological voice phenomena and voice disorders [6–9]. Recently, a pitch rise paradigm was investigated using a high-speed video system. In a single subject, both laryngeal and acoustic parameters were simultaneously determined [10].

The purpose of this study was to investigate the clinical application of the pitch rise task to examine the vibratory characteristics of vocal folds in young female adults.

Methods

Subjects

Twenty-nine female adults were examined. They were all students of Erlangen University Medical School, had a normal healthy voice and no history of voice disorder. Their mean age was 20.0 years, ranging from 17 to 36 years. Nine subjects indicated that they had had vocal training to a certain extent with choir experience or private singing lessons (at least 1 h/week). The remain-
ing 20 claimed to be untrained singers without any vocal training. None of the subjects was a professional singer or had a professional singing education.

Recording of Image and Acoustic Data

During phonation, the subjects were investigated with a rigid 90-degree endoscope (Wolf Corp., Germany), connected to a digital high-speed video camera and fed by a 250-watt xenon light. The camera sensor had a spatial resolution of 128 × 64 pixels (128 from left to right, 64 from dorsal to ventral) and a fixed time interval of 0.27 ms between the video frames, which corresponds to a recording rate of 3,704 frames/s. A recording time of 2 s was saved to mass storage. During the examination procedure, the subjects were sitting down in a chair and received topical anaesthesia, when necessary. They were examined twice and were asked to perform two different phonation tasks: (1) sustained phonation (‘sustain the vowel /e/, as in bed, at a comfortable pitch’) and (2) pitch rise (‘start phonation with an arbitrary low pitch and raise it as high as possible’).

No further instructions were given about the duration of the manoeuvre or the starting and final pitch.

High-speed video and acoustic signals were recorded in parallel. The acoustic signal was picked up with a condenser microphone (type B&K 4129, Bruel & Kjær, Copenhagen, Denmark) mounted on the top of the camera. The distance between the microphone and the mouth of the subject was approximately 20 cm. The acoustic signal was digitized at 44.4 kHz, 16 bits and saved on hard disc. Sound pressure levels (dB linear) were determined from the digital signals only. Calibration of the recording unit was performed as described in Huang et al. [11].

Evaluation of Image and Acoustic Data

Before analysis, high-speed recordings were processed with methods which had been described earlier [12]. Briefly, the evaluation algorithm performs fully automatic segmentation of the glottis within each frame (fig. 1). Additionally, it provides an estimation of the glottal midline by regression. Hence, it derives all important vibratory parameters as glottal area, vocal fold amplitudes and glottal length. Figure 1 shows 8 frames, covering 1 glottal cycle of 3.78 ms from a single subject, grabbed out from a high-speed sequence by omitting every second frame. The lower section of figure 1 shows the same images after the glottis segmentation and computation of the glottal midline.

The evaluation regards the vocal folds at their maximum amplitude perpendicular to the glottal midline, which is usually observed in the medial glottal third (50% from ventral to dorsal of the membranous part of the glottis). For computer memory reasons, only partial sequences instead of the complete recordings were used, each of them covering a time range of 17.3 ms (i.e. 64 frames).

For the sustained phonation condition, 1 partial sequence was evaluated. For the pitch rise condition, 3 partial segments (phase I, phase II and phase III) were depicted, representing the beginning of the ‘glissando’ (phase I), the middle (concerning the frequency) of the glissando (phase II) and the end of the glissando (phase III).

The type of glottal closure was judged by an experienced clinician and classified into 1 of 6 categories (1–6) according to Södersten and Lindestad [14] by visual inspection: (1) a 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 third of the vocal folds; (5) incomplete closure of the posterior two thirds of the folds, and (6) incomplete closure all along the vocal folds. Beyond this, 4 (irregular) types of incomplete spindle-shaped glottal closure patterns were taken into account.
Results

Sustained Phonation

Figure 3 shows the fundamental frequencies and sound pressure levels extracted from the acoustic recordings of all 29 subjects during sustained phonation. Subjects No. 1–20 belong to the untrained voice performers, while subjects 21–29 form the experienced singer group. Neither fundamental frequency nor sound pressure level differs significantly between the two groups.

The glottal closure types, subjectively classified as described above, showed not even 1 case of irregular closure types. The distribution of the regular closure types 1–6 is plotted in figure 2. Only 5 out of 29 subjects (17%) exhibited complete glottal closure. A ‘mean closure type’ of 2.7 ± 0.9 was calculated.

Fig. 1. a Eight single frames of a high-speed (HS) sequence representing 1 glottal cycle and covering a time interval of 3.78 ms. b Same as a with additional marking of the glottis and the glottal axis as derived by the image-processing procedure. The segmentation of the glottis forms the basis for further derivations of laryngeal parameters.
Pitch Rise

The parameters were determined from the acoustic signal, separately from the starting, intermediate and final phase of the pitch rise (index 1–3). The distribution is shown in figure 4a: the starting pitches $f_1$ range from 165 to 276 Hz, the intermediate pitches $f_2$ vary between 223 and 421 Hz, while the final pitches $f_3$ are within 263 and 724 Hz. By converting the frequency ranges into semitones, the mean frequency ranges for the untrained and trained group were 12.9 ($\pm$ 3.68) and 15.5 ($\pm$ 3.0) semitones, respectively. The difference was found to be significant using the Wilcoxon rank sum test ($p = 0.0209$).

During the pitch rise the test persons showed no consistent behaviour for the sound pressure level (fig. 4b). In the starting, intermediate and final phases, the mean sound pressure level was 67.2 $\pm$ 4.8, 75.9 $\pm$ 3.1 and 72.5 $\pm$ 7.2 dB, respectively. A
**Fig. 4.** Frequencies (a) and sound pressure levels (b) during the pitch rise manoeuvre (phase I = square, phase II = triangle, and phase III = circle) for all subjects.

**Fig. 5.** Distance of the vocal folds (in the mean third of the glottis) as a function of time. Phases I–III correspond to the starting (a, 198 Hz), intermediate (b, 410 Hz) and the final phases (c, 724 Hz) of the pitch rise. Results of a single subject (No. 6) derived by image processing.
level increase from the starting to the intermediate phase was observed in all but 1 of the subjects. From the intermediate to the final phase, only 14 subjects increased their voice sound level.

High-Speed Recordings
Figure 5 shows the extracted distance in the middle of the vocal folds as a function of time for a single subject separately for each phase (beginning, intermediate and end) of the pitch rise. The oscillations show substantial variations between the 3 phases. The starting phase was characterized by large vibratory amplitudes and comparatively long intervals of glottic closure, with a complete closure and an open quotient of approximately 0.6. Within the intermediate phase, fundamental period and vibratory amplitudes were both reduced by a factor of 2. During the entire cycle, the minimum distance between the vocal folds was not absolute, but almost zero, indicating a discretely incomplete glottal closure. The final phase was characterized by very small amplitudes of less than 3 pixels, a continuously open glottis and very short cycle durations. The vibratory amplitudes decreased with increasing frequency. Similar oscillation patterns were observed for all subjects. For each phase, the corresponding mean vocal fold amplitude was 14, 12 and 9 pixels.

Figure 6 shows vocal fold amplitudes as a function of the frequencies $f_1$, $f_2$ and $f_3$ for all pitch rise recordings. For each subject, the 3 phases – starting, intermediate and final phase of the pitch rise – are represented by a plus (+), a circle (o) and

![Figure 6. Vocal fold amplitude (in pixels) as a function of frequency during the pitch rise. + = Phase I; O = phase II; * = phase III.](image-url)
Fig. 7. Distribution of the glottal closure types in phase I (a), phase II (b) and phase III (c) of the pitch rise. The corresponding mean frequencies are 212, 341 and 474 Hz.
an asterisk (*), respectively. The mean fundamental frequency was 209, 338 and 472 Hz, respectively.

The distribution of the glottal closure types is given in figure 7. In phase I, 7 out of 29 subjects achieved a complete glottal closure (type 1) and the ‘average’ over all subjects is 2.6 ± 1.0. With increasing pitch the degree of closure diminished, which was characterized by a right shift of the distribution: in phase II (mean 3.9 ± 1.2) and phase III (mean 5.3 ± 1.2) no test person was found with a complete glottic closure (type 1). The final phase III was characterized by a missing glottal closure in most of the subjects (20 of 29). For all phases, no irregular glottal closure types were observed.

Discussion

The laryngeal mechanisms during a vocal pitch rise in a group of young healthy women have been endoscopically recorded in real time. The recordings were image processed to quantify vibratory amplitudes and frequencies, and subjectively judged to categorize types of glottal closure. In contrast to investigations on vocal fold vibrations of sustained phonation at different pitches [14–17], this study focused on vocal fold oscillations during a continuous pitch rise. Compared to sustained phonation, a pitch rise requires instantaneous variation and adjustments of a number of laryngeal muscles. Hence, pitch rise is a sensitive exercise for the neuromuscular feedback loop.

Acoustics

Most of the subjects start the pitch rise exercise at a frequency which is comparable to the sustained phonation. Obviously, this starting pitch seems to be comfortable, at least for the situation during examination with a rigid endoscope. Frequency ranges varied substantially between the subjects and were dependent on vocal training. Since no instructions concerning the range of the pitch rise were given, the achieved upper frequencies may be regarded as the maximum ‘just comfortable’ frequencies. They are below the maximum pitches expected from voice range profiles [18, 19]. The reduced upper frequencies are also a consequence of the uncomfortable examination situation with the endoscope.

Vocal Fold Amplitudes

Vibratory amplitudes were measured as the time-dependent distance between the free vocal fold edges and the glottal midline during a glottal cycle. An evaluation programme, developed for the ‘sustained phonation’ paradigm, turned out to be satisfactory to automatic amplitude extraction when the pitch rise was subdivided into 3 phases. As shown in figure 5, vocal fold amplitudes varied between 30 and 5 pixels, with a strong tendency to decrease with increasing frequency. These semi-quantitative values may be roughly converted into metric units: the focus of the zoom endoscope did not have to be adjusted for the individual subjects. Hence, a singular calibration measurement by using a split laser projection system [20] allows an approximate estimation: 1 pixel corresponds to 0.094 mm, and the amplitudes varied between 0.5 and 2.8 mm. Compared to recently published data of male subjects the vocal fold amplitudes were smaller by a factor of 2/3 [20].
Glottal Closure

In the presented data, there was a strong correlation between the type ('amount') of glottal closure and the fundamental frequency. The laryngeal conditions of the sustained phonation were similar to those of the starting phase I of the pitch rise paradigm. For both situations a high incidence of at least partly incomplete glottal closure was observed. The results are supported by previous stroboscopic studies in young female subjects [14–17, 21].

However, in previous studies no significant frequency dependence of the glottal closure was found [14, 16]. In contrast, the present results show a highly significant relationship between amount and type of glottal closure and different pitches. This difference may arise from the fact that previous investigations were carried out using lower frequencies in the high pitch scenarios [14, 15, 17], which may give rise to different laryngeal mechanisms for sustained phonation and the pitch rise paradigm.

Conclusion

For the first time, quantitative data for the laryngeal processes during pitch increase in a group of young healthy women were presented. These data provide information about short-time laryngeal processes and hence deeper insight into the biomechanical details of phonation. Further studies including normal male subjects and dysphonic patients are in progress.

Acknowledgements

The authors would like to thank Dr. Maria Schuster and Dr. Peter Kummer for performing the high-speed recordings and Dr. L.A. McKenna for her constructive suggestions and comments in preparing this manuscript. The work was supported by grants from the ELAN Fund of the University Hospital Erlangen and the German Research Foundation (DFG Ey15/10 and Ho 2177/2).

References