Detecting Inaudible Vocal Organ Changes Through Glottal Inverse Filtering

***,†Ahmed Geneid, *Marjo Rönkkö, ‡Risto Voutilainen, ‡Liisa Airaksinen, ‡Elina Toskala, §Paavo Alku, and *Erkki Vilkman,** **‡Helsinki, §Espoo, Finland, †Ismailia, Egypt*

Summary: The aim of this study was to investigate if there were objective quantities extracted from the speech pressure waveforms that underlay inaudible changes in the symptoms of the vocal organ. This was done through analyzing 180 voice samples obtained from nine subjects (five females and four males) before and after exposure to a placebo substance (lactose) and an organic dust substance. Acoustical analysis of the voice samples was achieved by using glottal inverse filtering. Results showed that the values of primary open quotient and primary speed quotient changed significantly ($P < 0.05$) as did the amplitude quotient ($P < 0.01$). Exposure to lactose resulted in significant changes of secondary open quotient ($P < 0.05$) but opposite to effects found for exposure to organic dust. Modeling of the vocal tract into cross-sectional planes revealed that the immediate plane above the vocal folds correlates inversely with the feeling that voice is tense, or feeling the need to make an effort when speaking in addition having a feeling of shortness of breath or the need to gasp for air. Such results may point to acoustically detected subclinical changes in the vocal organ that the subject him/herself feels while they remain perceptually undetected by others.

Key Words: Organic dust–Voice–Acoustics–Perceptual assessment–Inverse filtering–Vocal organ–Vocal tract–Allergy.

INTRODUCTION

Exposure to organic dust among susceptible subjects may cause significant changes in the vocal organ symptoms.[1] In an earlier study, we showed that a number of subjective vocal organ symptoms and signs change after exposure to organic dust. These symptoms and signs included a hoarse, husky, or tense voice that required extra effort when speaking and had a difficulty in starting phonation ($P < 0.05$). Other statistically significant symptoms that changed included a feeling of shortness of breath or the need to gasp for air and a feeling that the voice is weak or that it does not resonate ($P < 0.01$). However, voice clinicians were not able to detect changes in these characteristics upon perceptual assessment of subjects' voice samples.

Our hypothesis was that the vocal organ may react to exposure to organic dust in a way that maintains the perceived end product, voice, by changes within the organ itself. Such hypothesized changes may be subclinical, or undetectable by subjective perceptual assessment but still detectable by means of objective acoustic measures. Accordingly, the present study aims of investigating, if our hypothesis on compensation and/or changes in the vocal organ may be proved through significant changes in acoustical parameters.

As far as we know, acoustic correlates with changes in vocal organ symptoms that may be inaudible to the voice clinicians have not been studied before in medical research. In clinical practice, we have seen patients who complain of voice symptoms, which we as voice clinicians are not able to detect on a perceptual basis. This research investigates if those significant changes in vocal organ symptoms that we reported in the previous study[1] correlate with any of the glottal inverse filtering (IF) parameters. Thus IF in this study, serves as our acoustic method of evaluation especially in the presence of inaudible vocal organ changes and subclinical changes in voice symptoms. IF *per se* has been rarely used in such acoustic analyses contexts. This is despite the fact that IF constitutes a noninvasive method of estimating the source of voice, the glottal volume velocity waveform, and analysis of (potential) changes in it.

The IF is based on the source-filter model of human voice production that was proposed by Fant.[2] The model assumes that voice production can be modeled by three separate processes: the glottal flow, the vocal tract, and the lip radiation. Although the source-filter model is a simplification of the real human vocal apparatus, it can be considered accurate enough for most speech sounds especially the nonnasalized vowels. The specific IF method used in this study is the Iterative Adaptive IF (IAIF) method, which automatically estimates the glottal flow waveform from the acoustic pressure signals that are recorded in a free field. Therefore, no flow mask required to be worn.[3]

Analysis of voice production using IF usually comprises two stages. The first stage in which estimates of the glottal flow are computed. The second stage is parameterization, which involves quantifying the obtained waveforms with properly selected numerical values. The authors in this study used *TKK Analysis and Parameterization Toolkit (Aparat)*,[4] which is a voice source IF and parameterization software toolkit. Aparat is an open-source software developed in the department of Signal Processing and Acoustics, Aalto University, Espoo, Finland. It is available free of charge at: http://sourceforge.net/projects/aparat/. The software package has already been shown to be a useful tool and has been adopted by several speech research groups.[5–10]

Accepted for publication February 15, 2011.
Declaration of interest. The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.
From the *Department of Otolaryngology and Phoniatrics, Helsinki University Central Hospital, Helsinki, Finland; †Department of Otolaryngology, Suez Canal University Hospital, Ismailia, Egypt; ‡Finnish Institute of Occupational Health, Helsinki, Finland; and the §Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland.
Address correspondence and reprint requests to Ahmed Geneid, Department of Phoniatrics, Ear, Nose, and Throat Hospital, Helsinki University Central Hospital, PO Box 220, FI-00290 HUCH, Helsinki, Finland. E-mail: ahmed.geneid@hus.fi
Journal of Voice, Vol. 26, No. 2, pp. 154-163
0892-1997/$36.00

doi:10.1016/j.jvoice.2011.02.005

AIM OF THE STUDY

The aim of this study was to search for acoustic correlates to significant changes in some vocal organ symptoms using IF.

SUBJECTS

Nine volunteer subjects were studied (five females, four males). Volunteers ranged in age from 26 to 60 years (mean age 40.1 years). The subjects were referred to the Finnish Institute of Occupational Health for inhalation challenge tests[11] because of suspected occupational asthma or occupational rhinitis in reaction to certain organic dust substances in their working environments. Their occupations were: five bakery workers, one confectionery industry worker, one laboratory animal keeper, one carpenter, and one science teacher. None of the subjects had any vocal training. Subjects had neither a voice disorder nor a respiratory tract infection at the time of presentation for the challenge test.

METHODS

This is a prospective cohort study in which all subjects were exposed to two substances in a 6 m^3 ventilated airtight chamber. One of the substances was the specific organic dust that was suspected to cause occupational rhinitis or asthma to the individual subject and the other was a placebo (Lactose). Lactose and the organic dust substances were administered through air pulses to lactose or organic substance powder cup inside the ventilated exposure chamber. Lactose is widely used as placebo substance when testing for occupational asthma or rhinitis, without known allergic reactions. Exposure tests were done single blinded. Subjects were not aware to which substance they are going to be exposed, however, some of them were able to notice that they are being exposed to the organic substance based on the reactions they started to develop or the familiarity of the agent. Six subjects were exposed to rye, two to sawdust, and one to gum acacia. Because of the occupational diagnostic procedure of our patients, the occupational asthma tests, the exposure order needed to ensure that the asthma was in a reasonably steady state. Therefore, the first exposure was always lactose, and the organic dusts followed afterward on a later day.[12] The substances were delivered on different days, 1–6 days apart. The beginning of the exposure varied between 9.30 AM and 00.50 PM, starting on average at 10.16 AM. Each of the subjects was given a visual analog scale (VAS),[13] a continuous scale from 0 to 10, on which the subject self-assessed different voice parameters before and after exposure to lactose and the organic dust. All nine subjects were instructed to fill out a VAS within 5 minutes before and after either exposure. The scale included 12 voice and throat parameters which were: (1) My voice is overstrained, (2) My voice is hoarse or husky, (3) I have a feeling of a lump in my throat, (4) I feel like I have a choker around my neck, (5) I have a feeling of mucus in my throat and/or I need to clear my throat frequently, (6) My throat is dry and/or itchy, (7) My voice is weak/my voice does not resonate, (8) My voice is tense or I feel I must make an effort when speaking, (9) My voice is creaky, (10) My voice often breaks when I speak, (11) I feel short of breath/I need to gasp for air, and (12) I feel difficulty in starting phonation.

Voice samples

Voice recordings for each subject were made before and after each of the two exposures in a sound proof booth close to the airtight chamber in which subjects were exposed to lactose or the specific suspected organic dust substance. Voice recordings were always carried out within 5 minutes before and after the exposure.

The voice samples taken from each subject consisted of two stages. The first stage was a reading sample from a Finnish text consisting of three passages describing weather conditions. The text was repeated three times by each speaker. Each recitation took approximately 1 minute. The middle recitation was chosen for further analyses. The weather forecast text was specifically tailored to comprise several words with a long vowel [a] surrounded either by fricatives [s] or unvoiced plosives [k], [p], and [t] to obtain nonnasalized vowels of high first formant for the IF analysis. The second stage of the recordings involved vowel [a] sustained phonations on increasing vocal intensity. Four intensity categories were used that corresponded to sound pressure levels (SPLs) of 60, 70, 80, and 90 dBA. The SPLs were measured by portable SPL apparatus, which was placed on a table, 40 cm from the lips of the subjects. Each speaker was trained to follow the display, and to produce the sustained vowel with the given four SPL values. The production was repeated whenever needed to achieve best sustained phonations at the given values.

A total of 180 recordings of the speech pressure signals were made using a portable hard disk player (iRiver 140, Iriver Ltd., Irvine, CA) with a sampling rate of 44.1 kHz. A head-mounted condenser microphone (AKG, C444, AKG Acoustics GmbH, Austria) was located on one side of the mouth of the subject at a distance of 4 cm from his or her lips. This distance was always monitored by one of the authors (M.R.).

Voice samples were later transferred to a personal computer in which they were edited using audio editor (Adobe audition, Version 2.0 by Adobe Systems Incorporated, San Jose, CA). From the first part (reading sample), the word starting the second passage, "Kaatosade" (Finnish for "torrential rain"), was selected and the long [a] in the first syllable of this word was used in the IF analyses. For the second stage (vowel [a] sustained phonations), a segment of 50 ms was cut from the middle of the utterance and also used in the IF analyses.

Glottal IF

The audio files were then inverse filtered using *TKK Aparat* software.[4] The specific IF method used in this study was IAIF.[3] The IF parameters included both time-based parameters namely: fundamental frequency (F0), primary open quotient (OQ1), secondary open quotient (OQ2), open quotient from amplitude domain (OQa), quasi-open quotient, amplitude quotient (AQ), normalized AQ (NAQ), closing quotient, primary speed quotient (SQ1), and secondary speed quotient; in addition to the frequency-based parameters: Harmonic level difference, parabolic spectral parameter, and harmonic richness factor. The IF method is described in detail in the study by Fant[2] and Airas.[4] Calculation of various parameters called for determining several critical time-instants from the glottal flows and their