RELATING VOCAL FOLD AMPLITUDE OF VIBRATION TO SKIN ACCELERATION LEVEL ON THE ANTERIOR NECK

P. S. Popolo, I. R. Titze
Department of Speech Pathology and Audiology, University of Iowa, IA, USA and National Center for Voice and Speech, The Denver Center for the Performing Arts, Denver, CO, USA

Abstract: The purpose of this research was to determine if a relationship between vocal fold amplitude of vibration and skin acceleration level could be found using regression techniques. The effects of accelerometer location and phonation frequency were examined.

Keywords: vocal folds, vibration, acceleration

I. INTRODUCTION

The ability to measure amplitude of vocal fold vibration in vivo is of major importance in the field of speech science. In voice dosimetry, vocal fold vibration in human subjects during prolonged periods of speaking is studied in order to determine the effects of exposure to self-induced tissue vibration in vocalization [1]. Amplitude of vibration (A) is a variable in the calculation of two of the dose measures, distance dose (Dd) and energy dissipation dose (Dc). These measures are important for understanding vocal fatigue and recovery, especially among professionals who rely on their voice for their livelihood.

In the current voice dosimetry study being conducted at the NCVS, the doses are calculated from skin acceleration (SAL) measured at the jugular notch, on the anterior neck of the subjects. The derivation of A from SAL requires using a series of empirical equations based on previously published canine model and human subject data, and a calibration curve based on a lengthy data collection session in the laboratory with each dosimetry subject.

The purpose of this research was to determine an equation relating SAL to A using regression techniques, for predicting A from SAL. Seven different sites on the anterior neck were investigated. Human SAL and A data were obtained in vivo during standard laryngeal exams using custom equipment and state-of-the-art imaging and audio recording and processing techniques.

II. METHODS

A. Subjects, Materials, Tasks, and Data Collection

Two vocally healthy subjects, a male and a female, with no known vocal fold pathologies were administered videostroboscopic laryngeal exams with a rigid endoscope while wearing seven miniature accelerometers placed at various sites on the anterior neck, including at the jugular notch and above, below and lateral to the prominence of the thyroid cartilage. In order to obtain quantitative measurements of A in absolute dimensions, a two-point laser projection system was developed (Fig. 1a). The device projected two precisely-spaced green (wavelength = 532 nm) laser dots in the image frame, from which absolute dimensions could be determined in the laryngeal exam videos. Custom software was written to perform a frame-by-frame extraction of the absolute vocal fold length and glottal width at the midmembranous point, and A was calculated as half the width (assuming symmetrical displacement of the vocal fold edge from the glottal midline for stable, periodic vibration of normal, healthy vocal folds).

A lightweight, thin latex patch was designed to hold six accelerometers in a 2x3 array centered about the thyroid prominence, so that consistent, repeatable acceleration measurements could be made between the different subjects and different trials (Fig. 1b). The accelerometers were of the same type used in previous studies of long-term voice use [2], [3]. The patch was held firmly in place with a Velcro™ strap, and the surface of each accelerometer was attached to the skin with a temporary surgical adhesive. A seventh accelerometer was attached to the skin at the jugular notch with surgical adhesive and a small strip of medical tape. Fig. 2 shows the location of the seven accelerometer on the anterior neck. Fig. 3 shows a schematic of the experimental setup.

Subjects were asked to perform a series of sustained phonations on the vowel /i/ at a number of different intensity levels from soft to loud, and at three different pitches – comfortable, high, and falsetto. The accelerometer signals were amplified and digitally recorded to the hard drive of a data collection computer, at a sampling rate of 44.1 kHz. The video of the laryngeal exams was digitally recorded to the videostrobe host computer. All audio and video signals were time-synchronized so that SAL and A data points could be directly related to each other.

B. Data Processing and Statistical Analysis

Data was obtained from two separate trials for both subjects, with at least one week between trials. The subject/data sets were designated M01-1, M01-2, F01-1.
and F01-2. A cyclical plot of vocal fold amplitude of vibration \( A \) was extracted from the video signal for each data set, at a sampling rate equal to the video frame rate of 30 Hz. Strobe rate was set to Fast, yielding 1.5 glottal cycles per second and 20 frames per glottal cycle. Between 90 and 180 seconds of this cyclical representation was obtained for each data set. Root-mean square (RMS) values of \( A \) were obtained over 66.7 ms windows, overlapping by 33.3 ms (corresponding to a window size of one glottal cycle and an overlap of one-half cycle). The corresponding time segments of the SAL signals from all seven accelerometers were RMS-averaged with the same window duration and overlap, so the sequence of RMS values of \( A \) and SAL were still time-synchronized.

Scatter plots were generated of the time-synchronized RMS values of \( A \) vs. SAL for each of the seven accelerometer signals. \( A \) was calibrated to mm and SAL was calibrated to m/s². It was attempted to fit the data to a simple linear regression model,

\[
A_{\text{predicted}} = b_1 \cdot \text{SAL} + b_0 \tag{1}
\]

where \( b_1 \) and \( b_0 \) are the regression coefficients corresponding to the slope and intercept, respectively, of the regression line. The model was chosen based on the observation that, for sinusoidal vibration, the relation between displacement \( x \) and acceleration \( a \) is given by

\[
a_{\text{RMS}} = -\frac{\sigma^2}{\sqrt{2}} x_{\text{RMS}} \tag{2}
\]

where \( \sigma \) is the radian frequency of vibration. Statistical methods were employed to determine if there were significant differences between the fits obtained at the seven different locations, i.e., whether measuring the acceleration at different locations made any difference in the resulting fits; and if so, to determine which location showed the highest correlation to the vocal fold amplitude of vibration extracted from the video signal.

III. RESULTS

In plotting the RMS values of \( A \) vs. SAL, it was found that there was a clustering of the data points according to the fundamental frequencies of the phonations. Since pitch was not a variable but rather a parameter of the study (each subject did the same phonations at three different frequencies), it was decided to parameterize each plot of \( A \) vs. SAL by the frequency groupings Low, Medium and High. The linear regression fits were determined for each frequency group and each accelerometer location, as follows:

\[
A_{\text{predicted}} = b_1 \cdot \text{Location}_\text{All} \cdot \text{SAL}_{\text{Location}_\text{All}} + b_0 \cdot \text{Location}_\text{All} \tag{3}
\]

Model #2:

\[
A_{\text{Location}_\text{Low}, \text{predicted}} = b_1 \cdot \text{Location}_\text{Low} \cdot \text{SAL}_{\text{Location}_\text{Low}} + b_0 \cdot \text{Location}_\text{Low} \tag{4}
\]

\[
A_{\text{Location}_\text{Med}, \text{predicted}} = b_1 \cdot \text{Location}_\text{Med} \cdot \text{SAL}_{\text{Location}_\text{Med}} + b_0 \cdot \text{Location}_\text{Med} \tag{5}
\]

\[
A_{\text{Location}_\text{Hi}, \text{predicted}} = b_1 \cdot \text{Location}_\text{Hi} \cdot \text{SAL}_{\text{Location}_\text{Hi}} + b_0 \cdot \text{Location}_\text{Hi} \tag{6}
\]

where Model #1 is the fit for the data points of all frequencies combined, for a given accelerometer location, and Model #2 is the set of fits for the data points grouped according to frequency of phonation, either low, medium or high, for a given accelerometer location.

By fitting all subject data sets to the above models, \( b_i \) coefficients (slopes) that were significantly different from zero could be obtained for most, but not all of the accelerometer location/frequency group data points. The test of non-zero slope is statistically the same as the test that the correlation coefficient \( r \) of the regression model is not equal to zero; i.e. that the linear regression equation is a valid representation of the relation between SAL and \( A \). This was the consistently the case for all subject data sets for the low frequency data, at all accelerometer locations. Subject/set M01-1 also had non-zero \( b_i \)’s for the medium frequency data, and subject/set M01-2 had non-zero \( b_i \)’s for all three frequency groups. For this subject/set, statistical analyses showed that there were significant differences among the slopes of the fits for the three different frequencies at each accelerometer location, and that there were significant differences among the different locations. Furthermore, there was significant interaction between the effects of accelerometer location and frequency of phonation for this subject/set, in that there was a wider variation among the slopes of the different locations at low frequencies, but less variation among the slopes of different locations at medium and high frequencies of phonation. Looking only at the low frequency data for this subject/set, it was further found that certain pairings could be made, statistically, between the left and right counterparts of each location, which says that there is little difference between a left-right pair in the six locations around the thyroid prominence. Also, though there is not enough statistical evidence to distinguish between these six locations, taken as a group they are significantly different from the seventh location, the jugular notch.

Visual inspection of the \( b_i \) coefficients for the same subject in the two different trials showed no consistency, even though the repeatability of absolute measurements of vocal fold amplitude of vibration with the two-point laser projection system and videostroboscopy had been shown in an earlier study [4].

IV. DISCUSSION
The reason for the lack of intra-subject repeatability may have to do with the mechanism by which vocal fold vibration is transferred through tissue and measured as skin acceleration, and further investigation is needed. The lack of significant correlation between SAL and A at higher frequencies and in falsetto production may be due to the changes in vocal fold length, stiffness, and depth of vibration which characterize these types of phonation. The amplitude of vibration may not be adequately described by a linear model, and the “error” of the estimate may come not only from measurement error but also from the effects of unmeasured variables or un-included predictors, such as stiffness and length. Also, a two-dimensional measurement of horizontal amplitude of vibration does not describe the movement of tissue in the inferior-superior direction, which may contribute to the acceleration measured on the neck.

For the subject/set M01-2, the signals from the seven accelerometer locations all provided significant information for predicting the vocal fold amplitude of vibration. A principal component analysis may allow one to determine the relative amount that each signal contributes to the prediction, and if a subset of the signals can provide a reasonable estimate.

V. CONCLUSION

The current data set shows that there may be a significant correlation between SAL and A at lower phonation frequencies, i.e., at habitual speaking pitch. This relationship may hold if other parameters of vocal fold vibration, such as length and stiffness, are isolated or held constant. Further investigation is needed with more subjects and more repeated measures per subject.

REFERENCES