Technical considerations in the design of a wearable voice dosimeter

Article *in* The Journal of the Acoustical Society of America · November 2002

DOI: 10.1121/1.4779276

5 authors, including:

Peter Popolo
Digital Health Consultants
18 PUBLICATIONS **708** CITATIONS

Karen Rogge-Miller
4 PUBLICATIONS **55** CITATIONS

Jan G Svec
Palacký University Olomouc
141 PUBLICATIONS **3,537** CITATIONS

Ingo R Titze
University of Utah
424 PUBLICATIONS **21,266** CITATIONS

The National Center for Voice and Speech Online Technical Memo, No. 5, 2002, version 1.1

Rev. Feb, 2005

Technical Considerations in the Design of a Wearable Voice Dosimeter

Peter S. Popolo[1], Karen Rogge-Miller[1],
Jan G. Svec[1], Ingo R. Titze[1,2]

[1] National Center for Voice and Speech,
The Denver Center for the Performing Arts, 1245 Champa Street, Denver, CO 80204, USA,

[2] Department of Speech Pathology and Audiology
The University of Iowa, 330-WJSHC, Iowa City, IA 52242, USA

This paper deals with the technical issues and requirements involved in the design of a voice dosimeter, a wearable device to be used by teachers to measure vocal dose on the job, at home and elsewhere during the total waking hours of each day. The concept of vocal dose is introduced and definitions of various dose measures are presented. Practical issues of the work in progress are discussed. These are related to the software design, including the extraction of basic speech parameters, data reduction and storage, user interface design and power management; and hardware considerations concerning microphone selection and attachment to the subject. Unresolved issues are mentioned, and possible solutions are discussed.

This memo can be downloaded at http://www.ncvs.org/ncvs/library/tech .

Keywords: NCVS Dosimeter, voice accumulation, technical design.

Note: This memo was originally written as a submitted abstract to the 7th International Conference on Spoken Language Processing (ICSLP) held in Denver, 2002. While more recent publications is available concerning the NCVS dosimeter (See Popolo, et al, JSLHR, 2005, NCVS Technical Memo No. 1, 2, and 4), this memo served an important part in the development of the NCVS Dosimeter. Therefore we decided to make it available to the public even if it meant doing so in a retrospective manner. The title page (current page) and the final page are new, while everything in between are from the original document. Also note that the Wilber James Gould Voice Center has been renamed "National Center for Voice and Speech" since the time of the original manuscript. -Eric J. Hunter, February 2005.

TECHNICAL CONSIDERATIONS IN THE DESIGN OF A WEARABLE VOICE DOSIMETER

Peter S. Popolo[1], Karen Rogge-Miller[1], Jan G. Svec[1], Ingo R. Titze[1,2,3]

[1]Wilbur James Gould Voice Center
The Denver Center for the Performing Arts
ppopolo@dcpa.org, kamill7@netscape.net, jsvec@dcpa.org, ingo-titze@dcpa.org

[2]Department of Speech Pathology and Audiology
University of Iowa
ingo-titze@uiowa.edu

[3]National Center for Voice and Speech

ABSTRACT

This paper deals with the technical issues and requirements involved in the design of a voice dosimeter, a wearable device to be used by teachers to measure vocal dose on the job, at home and elsewhere during the total waking hours of each day. The concept of vocal dose is introduced and definitions of various dose measures are presented. Practical issues of the work in progress are discussed. These are related to the software design, including the extraction of basic speech parameters, data reduction and storage, user interface design and power management; and hardware considerations concerning microphone selection and attachment to the subject. Unresolved issues are mentioned, and possible solutions are discussed.

1. INTRODUCTION

In the vocally demanding profession of teaching, excessive use of the voice can lead to the relatively frequent occurrence of voice problems. There is a need to measure the amount of voice use among teachers and monitor the occurrence of voice problems in order to establish occupational safety limits. Measuring devices known as "voice accumulators" have been developed in the past to measure voicing time (i.e., vibration time of the vocal folds) [1], [2], or the combination of fundamental frequency and voicing time [3], or intensity and voicing time [4] for periods of up to 12 hours. None of these devices, however, have been made commercially available.

Work is currently being done to design a "voice dosimeter" that will recognize voiced segments of speech from a single microphone worn by the subject, compute the voicing time, intensity (measured as sound pressure level, or *SPL*) and fundamental frequency (F_0) from the microphone signal, and store a time-synchronized record of these parameters over the course of a day. Although this work is still in progress, some design issues and requirements have been clearly defined and are discussed in this paper.

2. DEFINITION OF DOSE MEASURES

The first step in developing a voice dosimeter is to specify the parameters to be measured that quantify vocal load. The trauma to vocal fold tissue caused by loud or prolonged vocalization can be thought of as an exposure problem, and exposure over time is generally quantified by dose. Various possible vocal dose measures have been defined in previous research [5] that quantify different voice exposure factors. These are *time dose*, *vocal loading index*, *distance dose* and *dissipated energy dose*.

Time dose is a measure of the total time (in seconds) during which the vocal folds are vibrating, and is defined as

$$D_t = \int_0^{t_p} k_v \, dt \tag{1}$$

where t_p is the performance time (total period of time being recorded) and k_v is the voicing unit step function

$$k_v = \begin{cases} 1 & for \quad voicing \\ 0 & for \quad non-voicing \end{cases} \tag{2}$$

i.e., an indicator of when the vocal folds are actually vibrating. This function can be thought of as a switch or, in digital logic terms, a "gate".

A *vocal loading index*, first defined in [6], measures the total number of cycles of vocal fold vibration. It is expressed in kilocycles (thousands of cycles) and can be defined as

$$VLI = voicing\,time \bullet (F_0/1000) \tag{3}$$

where voicing time is as defined in Equation (1).

Distance dose measures the distance in meters that the vocal folds travel in continuous vibration, defined as

$$D_d = \int_0^{t_p} k_v 4AF_0 \, dt \tag{4}$$

where A is the vocal fold vibration amplitude, with $4A$ being the distance traveled by the vocal fold tissue in one vibratory cycle.

Energy dissipation dose is a measure of the energy dissipated as heat in the vocal folds during vibration, and is defined in joules/m^3 as

$$D_e = \int_0^{t_p} k_v p_d dt = \frac{1}{2}\int_0^{t_p} k_v \eta (A/T)^2 \omega^2 dt \qquad (5)$$

where η is the viscosity of the vocal fold tissue, T is the thickness of the vibrating part of the vocal folds, and $\omega = 2\pi F_0$ is the angular frequency in radians/second. The variables A, η and T can be derived from SPL and F_0 using empirical relations, as described in [5].

Thus it is clear that the dosimeter is required to register a time-synchronized record of the basic speech parameters k_v, SPL and F_0 in order to calculate dose as quantified by any of the above measures.

3. DOSIMETER DESIGN ISSUES

Generally the dosimeter can be implemented either as a custom-built hardware device, or as a "virtual" device in software on a commercially available portable computer. For our purposes, the latter approach was chosen, and it was decided to use a Pocket PC as the platform. Such a device has the advantages of a powerful, dedicated processor, large memory capacity and a 32-bit operating system based on the familiar Windows environment, with functions like audio input and synchronized communication to a desktop computer already present. Also, as the prices of Pocket PCs become more competitive, the cost becomes comparable to that of building a custom device, and there is less danger of obsolescence than with custom hardware. The Compaq iPAQ 3635 was chosen for the initial development based on its processing power and memory capacity.

Design of the dosimeter was thus reduced to software program development for the Pocket PC, with the main hardware considerations limited to selecting an external microphone, connecting it to the Pocket PC, and devising a comfortable, reliable and repeatable means of attaching the microphone to the subject.

3.1. Program description and flow

Software development was done using Microsoft® embedded Visual C++ for the Windows CE environment. The program was designed to calculate SPL, F_0 and voicing time using simple methods, and to implement an efficient scheme for data reduction and storage. In addition, an interactive portion was developed that prompts the user every two hours to verify the data stored during the previous two hours, and to perform a vocal effort and soft phonation test.

3.1.1. Calculation of voicing time, SPL and F_0

A flow chart of the current dosimeter program is shown in Figure 1. In order to process the speech signal from the input microphone, it needs to be divided into analysis windows, or frames. A frame duration of 30 ms was chosen for our application, since longer durations might not detect inter-syllabic pauses in typical speech. The corresponding frame rate was 33⅓ frames per second, or 2000 frames per minute. The range of sample values was normalized to [-1,+1] volts, and the root-mean-square (RMS) voltage of the sampled speech data

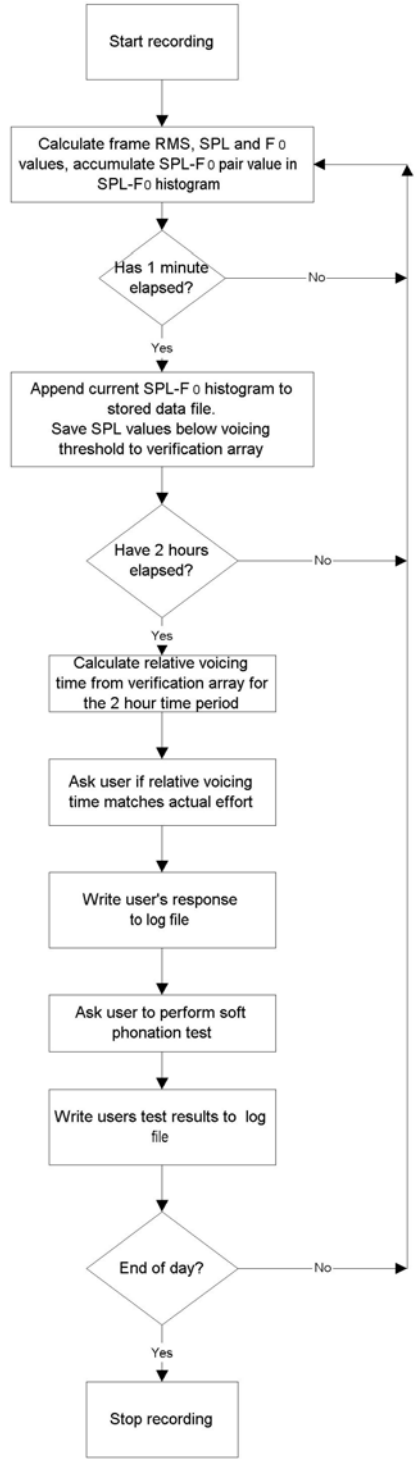

Figure 1: Dosimeter program flow chart.

was computed for each frame, which was in turn used to calculate the SPL value for the frame. A Brüel & Kjær 2238 sound level meter was used to calibrate the SPL, at a distance of 50 cm and with a linear weighting scale.

Voicing time was computed from the *SPL* data by setting a threshold level at 60 dB at 50 cm, below which it was assumed that the subject was not speaking. Thus the frames with *SPL* values above the threshold were assumed to be voiced and the others were assumed to be unvoiced.

F_0 extraction of the sampled data signal was accomplished by a cepstrum (short-term) analysis of each 30 ms frame. This method of analysis was chosen for its simplicity and reliability [7]. Processing time was kept to a minimum with the use of the FFT and inverse FFT algorithms. After the basic extraction, a post-processing algorithm was applied to correct for subharmonic tracking at half the fundamental frequency, the most frequent error to occur in this type of analysis [7]. The F_0 data obtained by this frequency extraction method has been compared to the data presented in [5], with favorable results.

3.1.2. Data reduction and storage

In order to reduce the amount of data storage required, the *SPL* and F_0 data pairs were accumulated into a two-dimensional histogram (*SPL* x F_0) each minute, and each histogram was appended to a data storage file throughout the day. The desired measurement range of *SPL* data was 54 dB, with a resolution of 2 dB, and the desired range of frequency data was 50-1000 Hz, or roughly 4½ octaves, in semi-tone steps. Table 1 shows a sample scheme for the *SPL*-F_0 histogram bin numbers and levels, in which the variables $h(i,j)$ represent the histogram levels in each bin.

A sample histogram containing approximately 7000 *SPL*-F_0 pairs is shown in Figure 2. The data was obtained from a laboratory recording of a female subject reading text from a children's book in exaggerated speech, much as a teacher might read aloud to a class of small children. The subject's mean *SPL* and F_0 over the entire reading were 68.3 dB at 50 cm and 216 Hz, respectively, corresponding to *SPL* bin number 6 and F_0 bin number 27, respectively. Note that the large number of values in bin *(1,1)*, corresponding to *SPL* < 60 dB at 50 cm and no pitch, represent the unvoiced frames of the speech signal. The rest of the levels correspond to the distribution of *SPL* and F_0 levels in the subject's recorded speech.

Since the number accumulated in each bin of the histogram will always be less than 2000 (the number of frames in one minute), it can be represented as a 2-byte unsigned integer. The size of each histogram is thus 3.2 KB. Based on a typical day with 14 hours of total waking time, the daily storage requirement is about 2.7 MB. This is a 5:1 reduction over storing 2000 *SPL* and F_0 values per minute as 4-byte floating decimal point variables for the same amount of time, which would require approximately 13.4 MB.

SPL Bin No.	F_0 Bin No.	1	2	3	...	55
	Pitch & F_0 (Hz) / *SPL* dB at 50cm	No Pitch	G1-G1# 49-51.9	G1#-A1 51.9-55	...	B5-C6 988-1047
1	< 60	*h(1,1)*	*h(1,2)*	*h(1,3)*	...	*h(1,55)*
2	60 - 62	*h(2,1)*	*h(2,2)*	*h(2,3)*	...	*h(2,55)*
3	62 - 64	*h(3,1)*	*h(3,2)*	*h(3,3)*	...	*h(3,55)*
...	...	...	...	...	...	...
28	112 - 114	*h(28,1)*	*h(28,2)*	*h(28,3)*	...	*h(28,55)*
29	> 114	*h(29,1)*	*h(29,2)*	*h(29,3)*	...	*h(29,55)*

Table 1: A sample scheme for *SPL*-F_0 histogram bins.

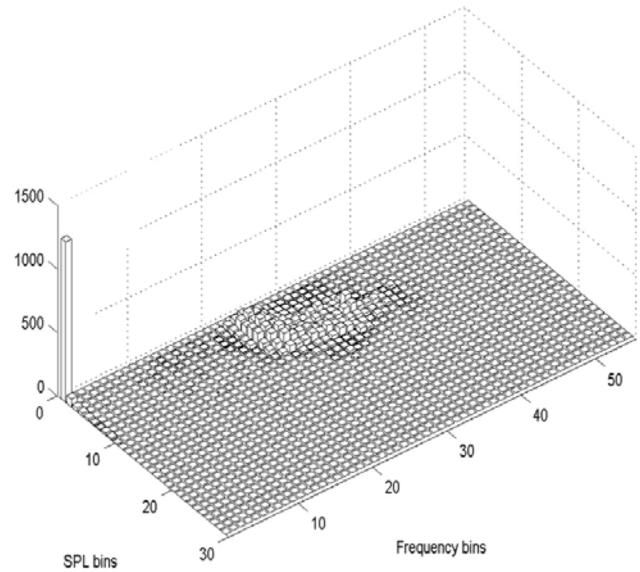

Figure 2: Example of *SPL*-F_0 histogram of speech from a female subject.

For a Pocket PC with 32 MB of ROM, the data storage required for a typical day is less than 10% of the total available storage space. The previous day's data file can be uploaded to a desktop PC on a daily basis, using the synchronized communication port provided on the Pocket PC's cradle, and emailed to a member of the investigating team. Once received, the files can be deleted from the Pocket PC's memory.

3.1.3. User interface design

While the dosimeter application is running, the main screen displays the program run-time as shown in Figure 3a. Every two hours, a verification screen is displayed which presents the user with their relative voicing time for the last two hours, calculated from the stored histograms. The graphical interface displays this number as a percent along with a textual interpretation, and asks the user if they agree or disagree. An example of this dialogue box is shown in Figure 3b.

The effort and soft phonation tests consist of four basic phonatory tasks that have been shown to be statistically significant indicators of vocal fold swelling [8]. A graphical interface prompts the user to perform the phonation tasks and press on-screen buttons corresponding to their numerical ratings of effort and soft-voce production. These tasks assume that the users have been trained to listen for specific signs of vocal deterioration in order to rate their performance. Examples of the dialogue boxes for the effort test and soft phonation test are shown in Figures 3c and 3d, respectively.

Since vocal use is not limited to only working hours, data needs to be collected for the teachers' total waking time. Currently, the program is started by the user each day and runs without the possibility of interruption for a fixed length of time. One issue still to be resolved is how the program will deal with the variable number of hours in a day, for instance on days

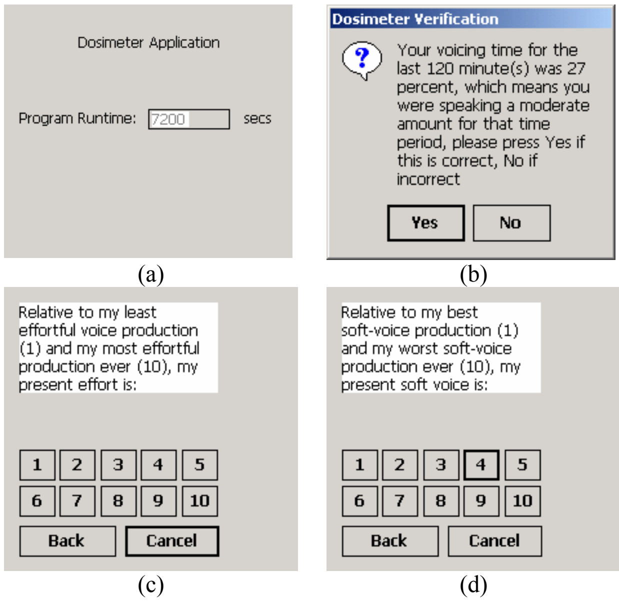

Figure 3: Examples dosimeter screens on the Pocket PC: (a) main screen, (b) user verification, (c) effort test and (d) soft phonation test.

when the teacher may rise earlier or later, or retire earlier or later than usual. A possible solution is to allow the user to end the program each day as well, but this raises the problem of inadvertent loss of data.

3.1.4. Power management

Under the current design, the Pocket PC operates for just under 4 hours when fully charged overnight. A future version of the dosimeter program is planned that will disable the display while the application is running and re-enable it when user interaction is required for the verification, effort and soft phonation tests. This is expected to greatly increase the runtime on the internal battery, and an external battery pack will be used as an emergency backup. When the external battery pack is depleted, the Pocket PC will revert to the internal battery. Once the battery power level reaches "low" (as indicated by the device's internal power monitor), an audible alarm will sound and a message will be displayed asking the user to replace the batteries in the external battery pack. The alarm will sound every 10 minutes until external power is restored.

3.2. Microphone considerations

The requirement for a single microphone is based both on practical considerations and a desire to minimize discomfort to the teachers. Because *SPL* measurement requires a known, fixed mouth-to-microphone distance, the built-in microphone on the Pocket PC cannot be used since it is difficult to control this distance during the teachers' various activities throughout the day. Also, the microphone must be able to discriminate between the users voice and background speakers or noise. Preliminary tests were done to connect a head-mounted microphone to the Pocket PC by breaking the connection to the internal microphone and running the wires out to an external connector, with good results. However, this type of microphone was found to be uncomfortable to the user for extended periods of time, and moved considerably during the day. Future work will consider a contact microphone attached to the surface of the skin with a Band-Aid (TM) at the location of the larynx. An unresolved issue is whether both *SPL* and F_0 can be accurately extracted from a microphone signal of this type.

4. CONCLUSIONS

Design of a device to measure the speech parameters related to vocal loading, known as a "voice accumulator", has historically been a difficult problem. A number of investigators have developed various prototypes since 1983, but no such device is commercially available to date. New research has been done to define various measures of vocal dose, which quantify vocal loading over time. Some technical issues have been defined in the current development of a new device, called a "voice dosimeter", which measures voicing time, *SPL* and F_0 in order to calculate dose and derive occupational safety limits for the high-risk profession of teaching. More work needs to be done before this device can be tested in the field.

5. ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health, grant number DC RO1 04224-01.

The authors would like to thank Russell Long for his participation in the early phases of this project, Sten Ternström for the suggestion of implementing the dosimeter on a Pocket PC, and Christian Herbst for his initial work on coding the soft-phonation test.

6. REFERENCES

[1] Watanabe, H., Shin, T., Oda, M., Fuaura, J., Komiyama, S., "Measurement of Total Actual Speaking Time in a Patient with Spastic Dysphonia", *Folia Phoniatr,* 39:65-70, 1987.

[2] Ryu, W., Komiyama, S., Kannae, S., Watanabe, H., "A newly devised speech accumulator", *Journal for Oto-Rhino-Laryngology and its Related Specialties,* 45:108-111, 1983.

[3] Ohlsson, A., Brink, O., Löfqvist, A., "A Voice Accumulator Validation and Application", *Journal of Speech, Language, and Hearing Research,* 32:451-457, 1989.

[4] Buekers, R., Bierens, E., Kingma, H., Marres, E. H. M. A., "Vocal Loads as Measured by the Voice Accumulator", *Folia Phoniatr Logop,* 47:252-261, 1995.

[5] Titze, I. R., Svec, J. G., Popolo, P. S., "Vocal Dose Measures", *submitted to Journal of Speech, Language, and Hearing Research*

[6] Rantala, L., Vilkman, E., "Relationship between subjective voice complaints and acoustic parameters in female teachers' voices", Journal of Voice, 13:484-495, 1999.

[7] Hess, W., Pitch Determination of Speech Signals, Springer-Verlag, Berlin, 1983.

[8] Bastian, R. W., Keidar, A., Verdolini-Marston, K., "Simple Vocal Tasks for Detecting Vocal Fold Swelling", *Journal of Voice,* 4(2):172-183, 1990

Acknowledgements

This work was supported by NIDCD grant No DC04347 from National Institutes of Health. Thanks to the Denver Center for the Performing Arts and their continued support as the NCVS Denver site host. Thanks also to other members of our team at both Denver and the NCVS University of Iowa Site.

Revisions

1.0 Peter S. Popolo, original abstract (2002)
1.1 Eric J. Hunter, formatted at a technical memo (Feb. 2005)