

The Goldilocks Passage and Scripts for Frequency Extraction, Voicing Detection, SPL Calculation and Vocal Dose Determination in Speech

Jan G. Švec¹, Peter S. Popolo¹ and 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 memo serves as an addition to two previous papers written by the above mentioned authors: 1) Titze IR, Švec JG, Popolo PS: Vocal Dose Measures: Quantifying Accumulated Vibration Exposure in Vocal Fold Tissues. Journal of Speech, Language, and Hearing Research 46(4): 922-935 (2003) [6]” and 2) Švec JG, Popolo PS, Titze IR: Measurement of Vocal Doses In Speech: Experimental Procedure and Signal Processing. Logopedics Phoniatrics Vocology, 28(4): 181-192 (2003) [4]. The first part of the memo offers the full text of the “Goldilocks story”, the passage of which was read by experimental subjects when determining preliminary normative values of vocal doses in [6]. The second part offers the Matlab scripts used by the authors to determine the vocal doses. These scripts incorporate a peak-picking algorithm for extraction of the fundamental frequency (F_0) and voicing detection from an EGG signal of speech as well as an algorithm to calculate absolute sound pressure levels (SPL) from speech signal registered by a head-mounted microphone. These scripts follow the procedure described in detail in [4] and can be downloaded from the web at www.ncvs.org/ncvs/library/tech.

Keywords: vocal dose measurement, voice accumulation, voicing time, electroglottography, voicing detection, F_0 extraction, absolute SPL measurement

1. The Goldilocks Passage

The Goldilocks passage has traditionally been used by the researchers of the National Center for Voice and Speech for voice evaluation, when large variations in intensity, fundamental frequency and voice quality are desired [5]. The Goldilocks passage enables incorporating different voice registers, such as chest/modal (as a possible representation of the voice of the big “Papa Bear”), and falsetto (as a possible interpretation of the voice of the “Baby Bear”). The plot of the Goldilocks story is traditional; the author of this particular version of the story is unknown to the authors of this memo. Since the text is rather long, only the first part of the Goldilocks story was used in the paper by Titze, Švec, & Popolo [6] - the reading stopped at the place marked by the * sign. For the sake of completeness, and for future research purposes, the whole text of the Goldilocks story is provided below.

Švec, Popolo & Titze: Goldilocks Passage and Scripts for Experimental Vocal Dose Measurement**Goldilocks and the three bears:**

“Once upon a time there were three bears. They lived in a house in the woods. One of the bears was Papa Bear. One was Mama Bear. And one was Baby Bear.

Each of them had their own porridge bowl. Papa Bear had a great big bowl. Mama bear had a middle-sized bowl. And Baby Bear had a wee little bowl.

And they each had a chair to sit on. Papa Bear had a great big chair. Mama Bear had a middle-sized chair. And Baby Bear had a wee little chair.

Each bed had a bed to sleep in, too. Papa Bear had a great big bed. Mama Bear had a middle-sized bed. And Baby Bear had a wee little bed.

One morning the three bears cooked some porridge for breakfast. While it was cooling they went out into the woods in search of a sweet honey. After they left, a little girl named Goldilocks passed by their house. She knocked on the door. No one answered, so she went right in.

Goldilocks was rather hungry, so she decided to eat some of the porridge from Papa Bear’s great big bowl. But it was too hot for her. Then she tasted the porridge from Mama Bear’s middle-sized bowl. But it was too hot for her as well. And then she tasted the porridge from Baby Bear’s wee little bowl. It was just right. So, she ate it all up.*

Goldilocks felt tired, so she sat down in Papa Bear’s great big chair. But it was too hard for her. Then she sat down in Mama Bear’s middle-sized chair. But, it was too soft for her. And then she sat down in Baby Bear’s wee little chair. It was just right for her, but it broke!

Then Goldilocks went upstairs to the three bears’ bedroom. First she lay down on Papa Bear’s great big bed. But it was too hard for her. Then she lay down on Mama Bear’s middle-sized bed. But it was too soft for her. And then she lay down on Baby Bear’s wee little bed. It wasn’t too hard and it wasn’t too soft. It was just right for her. So she pulled up the covers and went to sleep.

Soon after, the three bears came home. They noticed at once that something was wrong. “Somebody has been eating my porridge,” said Papa Bear in his great big voice. “Somebody has been eating my porridge,” said Mama Bear in her middle-sized voice. “And somebody has been eating my porridge and has eaten it all up!” cried Baby Bear in his wee little voice.

The three bears knew somebody was in their house. “Somebody has been sitting in my chair!” said Papa Bear in his great big voice. “Somebody has been sitting in my chair!” said Mama Bear in her middle-sized voice. “And somebody has been sitting in my chair and has broken it!” cried Baby Bear in his wee little voice.

The three bears hurried upstairs to their bedroom. “Somebody has been lying in my bed!” said Papa Bear. “Somebody has been lying in my bed!” said Mama Bear. “And somebody has been lying in my bed – and there she is!” cried Baby Bear.

Goldilocks woke up. When she saw the three bears, she jumped out of bed, raced down the stairs, and ran out the front door. Goldilocks ran as fast as she could through the woods. The three bears called for her to come back. But Goldilocks ran and ran, and the three bears never saw her again.”

2. Data processing for the calculation of vocal doses:

The experimental method was described in detail by Švec, Popolo and Titze [4] and the reader is referred to that paper for explanation of the procedure for determination of the vocal doses. This memo describes and offers the scripts, which were used for processing of the speech signal. The scripts calculate five different vocal doses from a speech signal: the *Time Dose* D_t (total voicing time), the *Cycle Dose* D_c (total number of vocal fold oscillatory cycles), the *Distance Dose* D_d (total distance traveled by the vocal folds in an oscillatory path), the *Dissipated Energy Dose* D_e , (total amount of heat energy dissipated in the vocal folds), and the *Radiated Energy Dose* D_r , (total energy radiated out the mouth). All the doses are calculated from three parameters of speech: the fundamental frequency (F_0), the sound pressure level (SPL) at 50 cm distance and the voicing-unvoicing parameter k_v , which is defined as [4;6]:

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

The method uses an EGG signal of speech to detect individual oscillatory periods of the vocal folds, to determine the F_0 of speech and to decide whether the signal is voiced or unvoiced (k_v). The detected periods are then applied to the audio signal (recorded simultaneously with the EGG signal) registered by a head-mounted microphone and the SPL level is calculated for each cycle individually. The absolute SPL level at 50 cm is determined by means of a two-step calibration procedure [4] which relates the head-mounted microphone signal to the signal registered by a sound level meter. When the F_0 , SPL and k_v values for each cycle are found, the doses are calculated using the definitions and empirical rules given in the papers of [6] and [4].

2.1. Files for the analysis:

Three audio files are needed for the determination of the vocal doses when using the scripts. All the three files are expected to be saved in a standard stereo wav format.

Audio File 1: A steady calibration tone registered by a microphone of a sound level meter. The sound level meter should be set to either “C” or “linear” frequency weighting and “fast” time response is recommended [1]. The sound pressure level of the calibration tone measured directly by the sound level meter shall be known. The calibration tone should be stored in channel 2 of the wav file, channel 1 of this file is not used (Figure 1).

Audio File 2: Simultaneous recording of the head-microphone signal and the sound-level-meter (SLM) microphone signal when the subject is producing sustained /a/. The head-microphone should be ideally at ca. 5 cm distance from the lips slightly to the side and out of the air stream. The sound level meter microphone should be 50 cm away from the lips. It is recommended to use a head-mounted microphone with a flat frequency response (note - omnidirectional microphones are usually better in terms of the frequency characteristics than

Švec, Popolo & Titze: Goldilocks Passage and Scripts for Experimental Vocal Dose Measurement

directional microphones since they do not suffer from the so called proximity effect [2;3]). The part of the sustained /a/ that is most stable with respect to the *SPL* value (about 1-2 s duration) should be selected and saved in a form of a standard wav file. The head-mounted microphone shall be in channel 1, the SLM signal in channel 2 of the wav file (Figure 2).

Audio File 3: Recording of the speech intended for the determination of the vocal dose (e.g., the Goldilocks passage). The recording should contain two simultaneously recorded signals – a head-microphone signal and an EGG signal. The polarity of the EGG signal should be set so that increase in the signal value corresponds to increase in the contact area between the vocal folds. The head-mounted microphone should be in the same position as it was during recording of the File 2. The head-mounted microphone signal should be saved in channel 1, the EGG signal in channel 2 of the wav file (Figure 3).

File 3 is sufficient if one wants to perform only the F_0 extraction and voicing detection. Files 1 and 2 are required for accurate determination of the absolute *SPL* values of speech (in dB re 20 micropascals at 50 cm distance) from the head-mounted microphone signal.

2.2. The Matlab Scripts:

The processing of the speech signals in Files 1-3 is done using scripts written in Matlab version 6. The scripts are directly available at www.ncvs.org/ncvs/library/tech. These are:

FREQSPL.m – the core script which performs all the processing of the data, does F_0 extraction, *SPL* calibration and absolute *SPL* calculation, employs empirical rules to derive lung pressure, vocal fold thickness, vibration amplitude of the vocal folds, dissipated power, radiated power and derives the doses for the entire signal. The script calls the following functions:

framer.m, which divides signal into frames of specific time interval and of specific time advance;

rms.m, which calculates a root mean square value over specified frames;

smooth.m, which smoothes the signal over frames;

alternpeaksJ.m, *maxminmaxJ.m*, *diffmaxminJ.m*, and *freqcheckJ.m*, which are used for extraction of the fundamental frequency of speech from the EGG signal;

gaincal_c.m, which calibrates the head-microphone signal for absolute *SPL* values;

spl.m, which derives the *SPL* value for each vocal fold vibration cycle;

dosesJ.m, which employs the empirical rules [4, 6] and calculates the final doses;

iseven.m, *isodd.m*, which are used to recognize even and odd numbers;

peak.m, which detects peaks in a vector of data;

row.m, which returns a row vector from a column (or row) vector;

zerox.m, which is used to detect zero crossings.

Furthermore one more script is included:

FREQSPLtest.m – script performing the same calculations as the script *FREQSPL.m*. The difference from the script *FREQSPL.m* is that it also plots figures of the temporary results along the different stages of signal processing. The script was found useful for debugging the frequency extraction algorithm and used also for generating the figures for this memo.

All these files are packed together in the file *VocalDoseMeasuresScripts.zip*.

The National Center for Voice and Speech Online Technical Memo No.1, April 2003, version 1.4

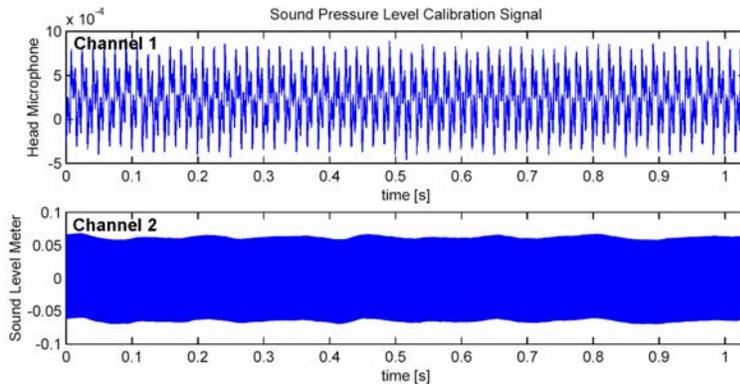


Figure 1: Signal for absolute SPL calibration as saved in the two channels of the wav file. The SPL calibration signal is in channel 2 (it contains a stable sound of 94 dB SPL produced by a calibrator which was attached to the sound level meter microphone), channel 1 is not used and contains only noise. (File *male_SPLcal.wav*).

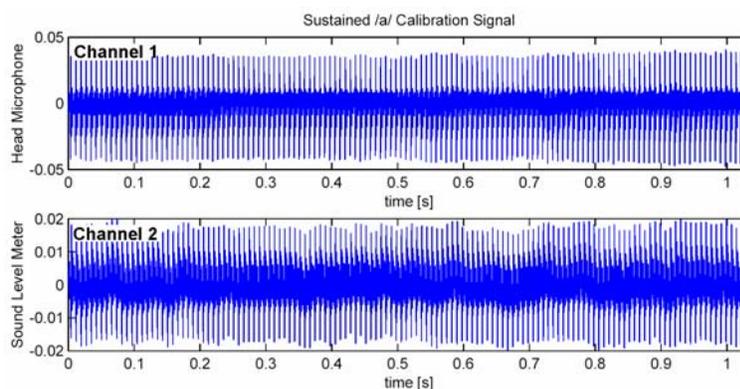


Figure 2: Sustained /a/ registered simultaneously with a head-mounted microphone and a sound level meter at 50 cm distance. Only the most stable segment of about 1 second duration was selected from the original phonation in order to allow automatic SPL calibration. The the head-microphone signal is in channel 1, the sound level meter signal is in channel 2). (File *male_sustain_a.wav*).

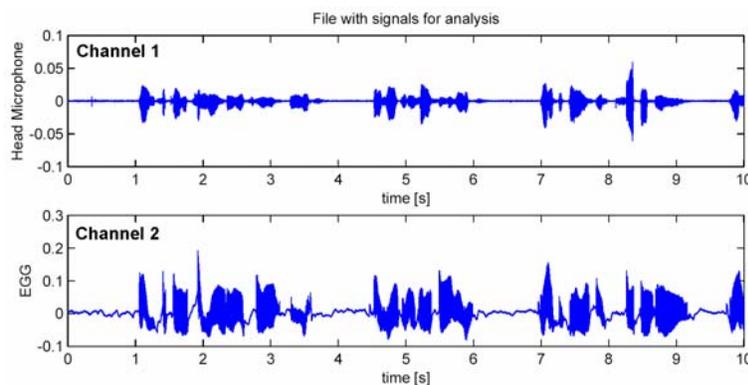


Figure 3: Display of the two channels of the wav file, which contains reading of the Goldilocks passage. Channel 1 contains the head microphone signal, channel 2 contains the EGG signal. (File *male_exag_reading.wav*).

2.3. Audio examples:

The zip file includes also examples of audio signals that can be used for the analysis. The files with the audio examples are:

male_SPLcal.wav/female_SPLcal.wav – two Files 1, containing the calibrator signal of 94 dB SPL for the male / female subject.

Švec, Popolo & Titze: Goldilocks Passage and Scripts for Experimental Vocal Dose Measurement

male_sustain_a.wav/female_sustain_a.wav – two Files 2, containing sustained /a/ vowel produced by the male / female subject, recorded simultaneously with the head-mounted microphone and the Sound Level Meter at 50 cm distance.

male_exag_reading.wav/female_exag_reading.wav – two Files 3, containing the head-mounted microphone signal and the EGG signal of a male/female subject reading 10 seconds of the Goldilocks passage (in exaggerated mode, as if reading for small children).

male_EGG_part.wav - another File 3, containing head microphone signal and EGG signal of a different male subject. This file was used for demonstrating the peak-picking procedure as shown in Figures 7-10.

2.4. Running the scripts:

In order to perform the complete calculation of the doses, one has to extract the scripts and wave files from the file *VocalDoseMeasuresScripts.zip* in one folder, open the Matlab program, set the working directory of the Matlab program to be the folder in which all the scripts and audio files are placed and run the script by typing “*freqspl*”.

The following message appears:

```
The script analyzes wav file which contains 2 channels recorded
simultaneously:
Channel 1 shall contain Head-Mounted Microphone signal of speech,
Channel 2 shall contain EGG signal of speech
.
Enter the name of the speech file or hit <Enter> for default
(default file: male_exag_reading.wav >
```

After entering the file name or hitting <Enter>, the user is asked to enter the low limit of the expected F_0 of the speech:

```
Enter the lower F0 limit in Hz and hit <Enter>
or hit <Enter> for default value 50 Hz >
```

If one is not sure about the value, for most of the voices, the default value of 50 was found to provide reasonably accurate results. (The value 50 can also be used to perform preliminary analysis after which a more specific value can be chosen.) Then, the user is asked to enter the upper F_0 limit of speech (F_{0upper}):

```
Enter the upper F0 limit in Hz and hit <Enter>
or hit <Enter> for default value 1000 Hz >
```

If the upper F_0 limit is not known, the value of 1000 was usually found to produce reasonably accurate results.

After entering the F_{0upper} value, this value is used to calculate the minimal duration of the oscillatory period T_{min} using the inverse relationship $T_{min} = 1/ F_{0upper}$. Then, the program reminds:

The National Center for Voice and Speech Online Technical Memo No.1, April 2003, version 1.4

Absolute SPL calculation requires two files:

- 1) File with SPL calibration signal from Sound Level Meter with known dB value
- 2) File with sustained /a/ registered with both the SLM and Head-Mounted Microphone

Both files shall have 2 channels:

- The Head Mounted Microphone signal shall be in Channel 1
- The Sound Level Meter signal shall be in Channel 2

.
If you do not have these files available,
the program can only perform F_0 extraction and voicing detection from the EGG
signal

.
Absolute SPL calculation?
Enter 1 for yes or 0 for no and hit <Enter>

At this point the user shall decide whether to perform only F_0 extraction or whether to do also the calibration for absolute *SPL* values. The F_0 extraction itself does not require the calibration for absolute *SPL* values. For dose calculations, however, the *SPL* calibration is required. After entering 1, the program asks for the *SPL* calibration file (i.e., the File 1 as described earlier):

Enter the name of the SPL calibration file and hit <Enter>
or hit <Enter> for default: male_SPLcal.wav >

After entering the name of the calibration file, the program asks for the sound level value measured by the sound level meter for the entered file:

Enter the SPL calibration value in dB for the SLM signal and hit <Enter>
or hit <Enter> for default value: 94 >

In cases of the two Files 1 provided as examples (*male_SPLcal.wav* and *female_SPLcal.wav*), the value was measured to be 94 dB, which is set as default.

Next, the program requires the name of the File 2, containing sustained /a/ registered simultaneously by the head-mounted microphone and the sound level meter microphone:

Enter the name of the file with sustained /a/ and hit <Enter>
or hit <Enter> for default: male_sustain_a.wav >

After that, the program asks whether to perform smoothing of the EGG signal. Smoothing decreases the noise level of the signal and makes the peak-picking detection to be less influenced by the noise. It is, however, not absolutely required. The smoothing procedure takes a running average of eight samples and was tested with signals sampled at 48000 Hz (as it is true for all the provided example files).

Do you want to smooth the EGG signal for smoother F_0 contour?
Enter 1 for yes or 0 for no and hit <Enter>

Švec, Popolo & Titze: Goldilocks Passage and Scripts for Experimental Vocal Dose Measurement

Depending on the computer used, the smoothing procedure can be time consuming, especially for signals with long durations. After finishing the smoothing, the program plays the audio waveform of the File 1. When playing the file, the program displays a figure of the complete EGG signal and asks for the noise value:

Input the noise threshold and hit the <Enter> key:

The noise value is important for distinguishing the voiced and unvoiced segments of speech. The user should determine it manually by inspecting the EGG signal in the displayed figure. In order to do it, the user shall use the “zoom in” tool to select and display a part of the signal without any voicing activity (Figure 4). When zoomed in, the peak-to-peak amplitude of the noise signal can approximately be determined (Figure 5). Baseline shifts are disregarded when determining the noise level. For the provided audio examples, the noise values were found to be, approximately:

male_exag_reading.wav – 1.2e-3 (smoothed), 2.8e-3 (unsmoothed)
female_exag_reading.wav - 2.2e-3 (smoothed), 4.5e-3 (unsmoothed)

The determined value shall be entered in the program and confirmed by hitting <Enter>.

Next, the program requests information on the maximal amplitude of the EGG signal during voicing by typing a message:

Input maximum peak-to-peak extent and hit the <Enter> key:

The same figure as used for determination of the noise level (Figure 4) can be used also to determine this value. The user should use the “zoom-out” tool to display again the complete EGG signal and then zoom-in to a part of the signal at which the amplitude is maximal (Figure 6). Again, the baseline shifts are disregarded when determining the value. The maximal peak-to-peak values of the EGG signal determined for the audio examples provided are, approximately:

male_exag_reading.wav – 0.16; *female_exag_reading.wav* - 0.12

The determined value shall be entered in the program and confirmed by hitting <Enter>.

On the basis of the entered values, the program performs peak picking of the EGG waveform and determines the individual oscillatory cycles of the vocal folds. The peak-picking algorithm has numerous phases, see the paper [4] for detailed description. At first, the program looks for all the maxima and minima of the captured waveform. Then, the global maxima and minima which are separated by more than T_{min} time interval (e.g., 0.001 s for the upper frequency limit of 1000 Hz) are selected from the maxima and minima. The number of selected maxima and minima is further reduced by applying a condition that there is only one maximum between two minima and only one minimum between two maxima. All these procedures are performed by the script “*alternpeaksJ.m*”. The program informs about this procedure by writing a message:

The National Center for Voice and Speech Online Technical Memo No.1, April 2003, version 1.4

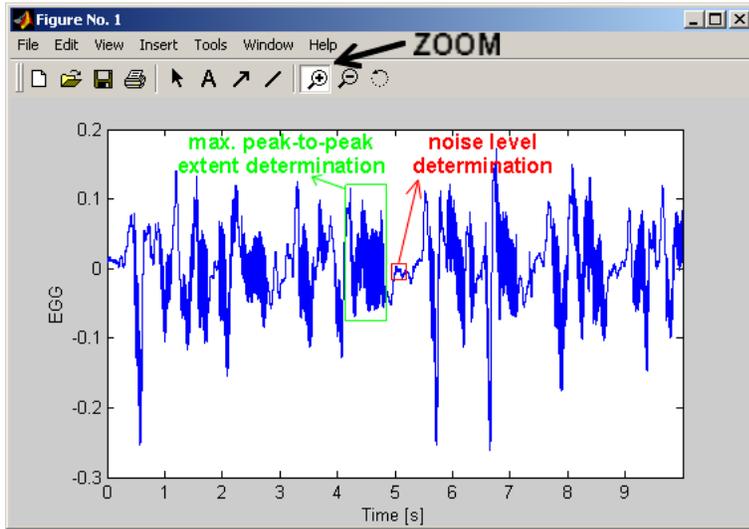


Figure 4: Example of the EGG signal waveform. For determination of the noise level a part of the signal with no voice activity is selected. For determination of the maximal peak-to-peak extent of the EGG waveform, the portion with greatest EGG amplitude is selected (the baseline shifts are disregarded). (File analyzed: *female_exag_reading.wav*).

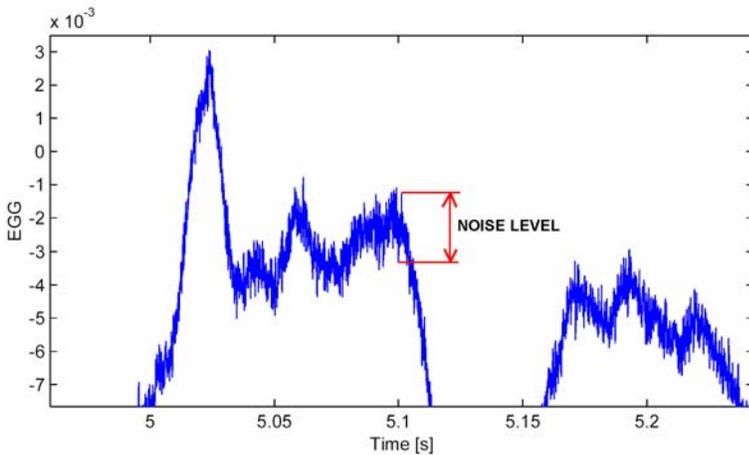


Figure 5: Empirical determination of the noise level from the unvoiced part of the EGG signal. The noise level is obtained by subtracting the minimum value from the maximal value (baseline shifts disregarded). In this example the noise level is $-0.0012 - (-0.0034) = 0.0022$ (relative units). (File analyzed: *female_exag_reading.wav*).

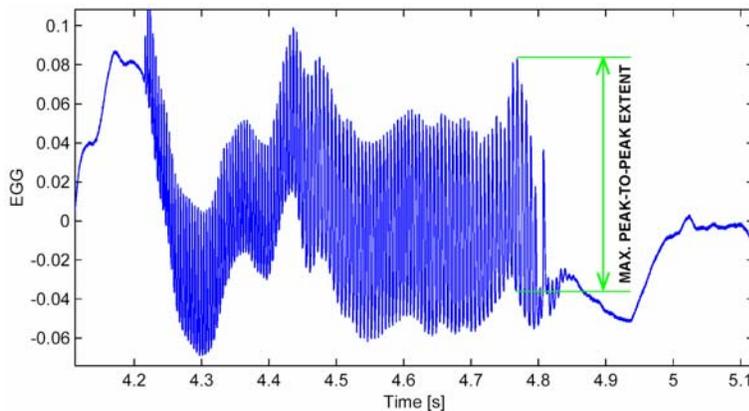


Figure 6: Empirical determination of the maximal peak-to-peak extent of the EGG signal. The extent is obtained by subtracting the minimum value from the maximal value (baseline shifts disregarded). In this example the peak-to-peak extent is $0.085 - (-0.035) = 0.12$ (relative units). (File analyzed: *female_exag_reading.wav*).

Švec, Popolo & Titze: Goldilocks Passage and Scripts for Experimental Vocal Dose Measurement

```
running function "alternpeaksJ.m"
FINDS PEAKS DISTANT BY MORE THAN SELECTED TIME (low-pass filtering)
AND MAKES SURE THAT ONLY ONE MINIMUM IS BETWEEN TWO MAXIMA
AND ONLY ONE MAXIMUM IS BETWEEN TWO MINIMA
please wait...
```

Depending on the computer used, executing this Matlab script can take some time, especially for signals with long durations. Figure 7 displays all the peaks of the EGG waveform (locmax and locmin) and the reduced set of minima and maxima (newlocmax2 and newlocmin2) after the execution of the script *alternpeaksJ.m*.

Then, the program runs the script *maxminmaxJ.m*, about which it informs by writing a message:

```
running function "maxminmaxJ.m"
USES ALTERNATING MAXIMA AND MINIMA FROM EGG
MAKES SURE THAT TWO NEIGHBOURING MAX ARE BOTH LARGER THAN THE MIN IN BETWEEN
AND TWO NEIGHBOURING MIN ARE BOTH SMALLER THAN THE MAX IN BETWEEN
```

As the message says, at this phase the program further reduces the set of selected maxima and minima by applying a condition that a valid maximum shall be larger than the minima between which it lies and, similarly, a valid minimum shall be smaller than both the maxima between which it lies. The new reduced set of maxima and minima is shown in Figure 8 as newlocmax3 and newlocmin3.

Further, the program runs the script *diffmaxminJ.m*:

```
running function diffmaxminJ
USING ALTERNATING MAXIMA AND MINIMA FROM EGG
MAKES SURE THAT THE NEIGHBOURING MAX AND MIN ARE DISTANT BY MORE THAN SELECTED
TIME
AND THAT THE DIFFERENCE BETWEEN THE NEIGHBOURING MAX AND MIN IS ABOVE NOISE
LEVEL
ELIMINATES SECONDARY PEAKS WITHIN A CYCLE WHICH ARE SMALLER THAN THRESHOLD
```

At this phase, the program further reduces the selected set of maxima and minima by simultaneously applying two conditions: the neighbouring maxima and minima shall be distant by more than $0.01 T_{min}$ and the difference between levels of the neighbouring maxima and minima is larger than the noise value. This eliminates the peaks resulting from the noise and the remaining set of maxima and minima is assigned as newlocmax4 and newlocmin4 (Figure 9).

The final procedure of the peak-picking algorithm removes a secondary maximum within an EGG cycle, which has been observed in some normal subjects. Before starting the procedure, the programs asks for a threshold value of the secondary peaks within a cycle:

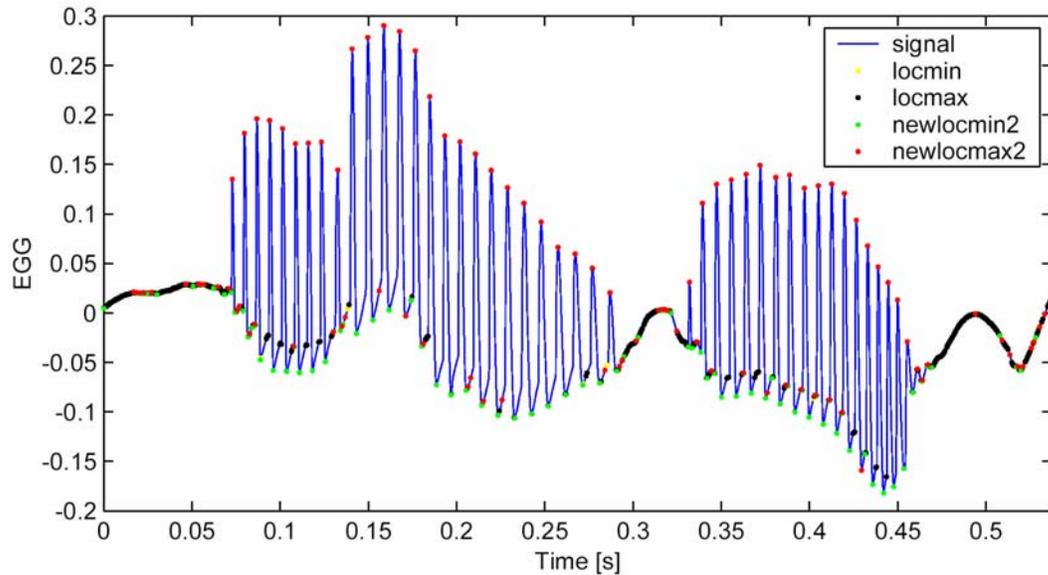
The National Center for Voice and Speech Online Technical Memo No.1, April 2003, version 1.4

Figure 7: All the local maxima and minima of the EGG waveform (black and yellow dots, marked as locmax and locmin in the figure legend) and the reduced set of maxima and minima after the execution of the script *alternpeaksJ.m* (red and green dots, marked as newlocmax2 and newlocmin2 in the figure legend). (File analyzed: *male_EGG_part.wav*).

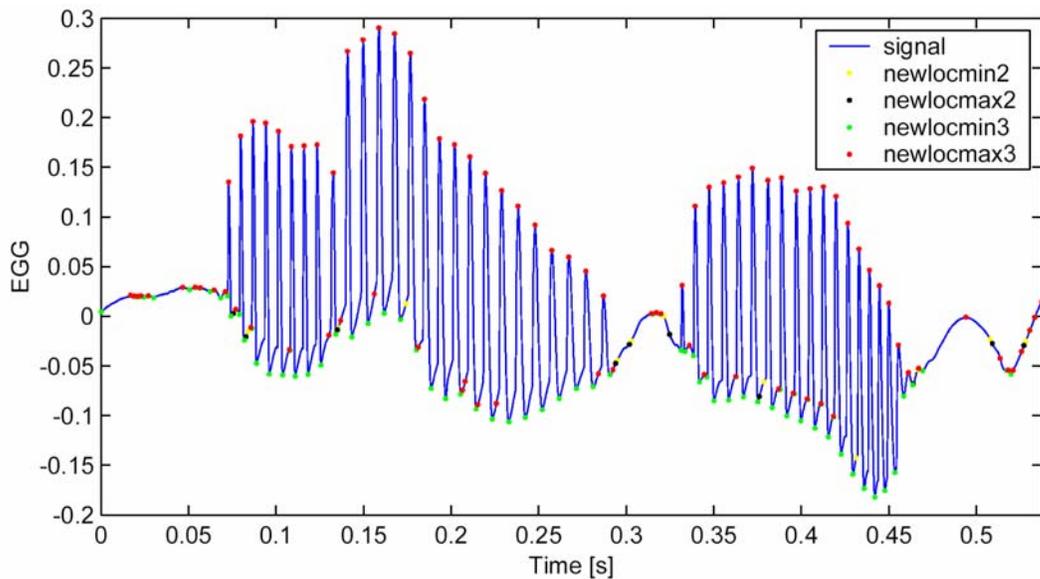


Figure 8: The maxima and minima of the EGG waveform after the execution of the script *maxminmaxJ.m* (red and green dots marked as newlocmax3 and newlocmin3 in the figure legend). The yellow and black dots are the maxima and minima, which were eliminated during step 7. (File analyzed: *male_EGG_part.wav*).

Švec, Popolo & Titze: Goldilocks Passage and Scripts for Experimental Vocal Dose Measurement

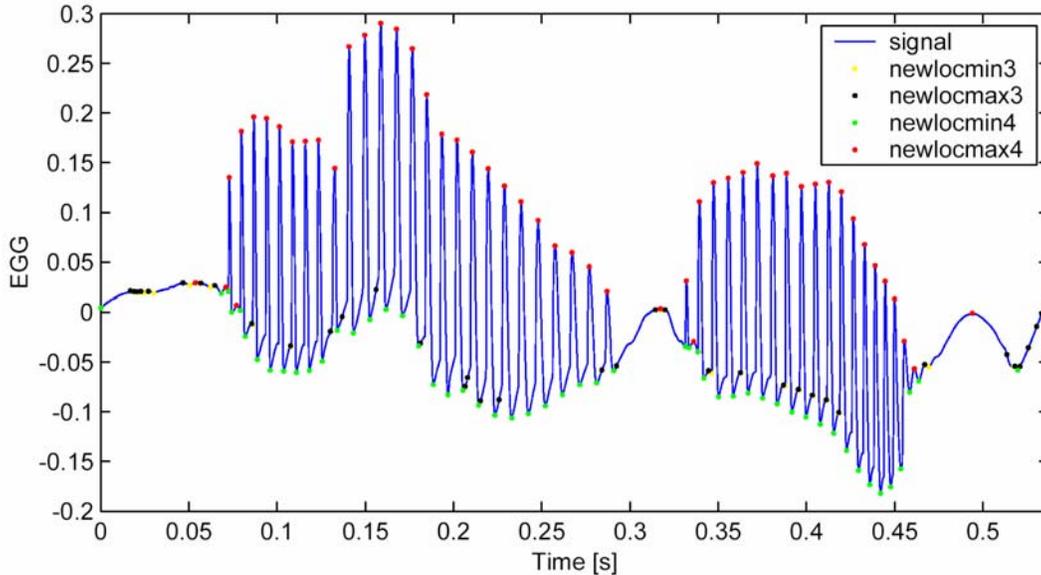


Figure 9: The maxima and minima of the EGG waveform after execution of the first part of the script *diffmaxminJ.m* (red and green dots marked as *newlocmax4* and *newlocmin4* in the figure legend). The yellow and black points are the maxima and minima eliminated during this phase. (File analyzed: *male_EGG_part.wav*).

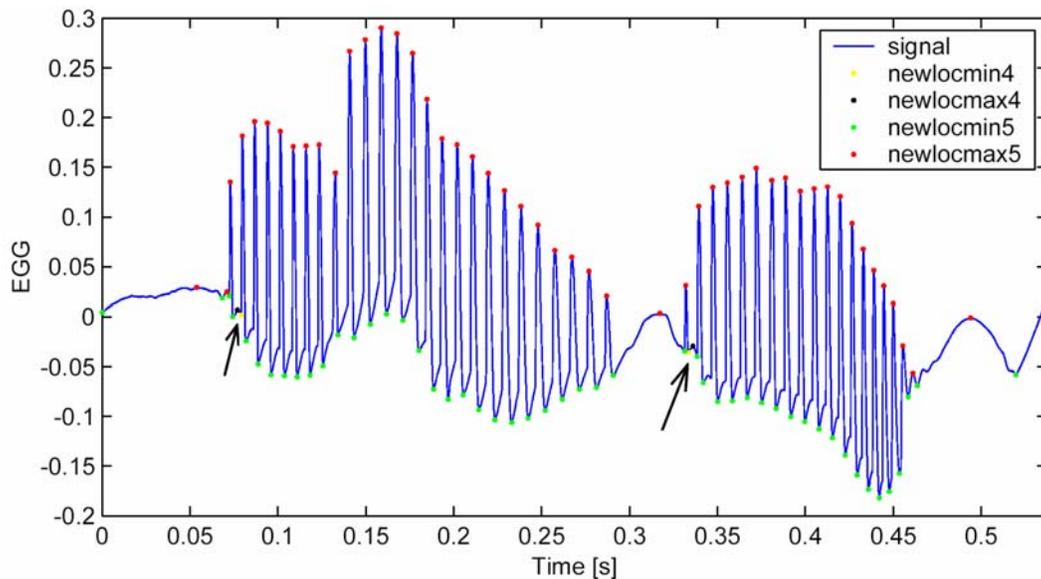


Figure 10: The final set of the maxima and minima (red and green dots, *newlocmax5* and *newlocmin5*). The arrows point to the secondary maxima and minima, which were eliminated by the second part of the script *diffmaxminJ.m*. (File analyzed: *male_EGG_part.wav*).

The National Center for Voice and Speech Online Technical Memo No.1, April 2003, version 1.4

Specify the threshold for secondary peaks within a cycle;
(e.g., for value 5, every secondary peak smaller
than 1/5 of the main peak will be eliminated).
CONDITION: EGG CONTACT MUST BE ORIENTED UPWARDS!
Input your value or hit <Enter> for default (5.5):

The default value of 5.5 was empirically found to work best for EGG signals without artifacts. In case of an unusual EGG waveform, with multiple peaks within a single oscillatory cycle (i.e., more than one secondary maximum within a single cycle), it is recommended to enter a large value (100 or greater) and use a noise threshold that is greater than the peak-to-peak value of the EGG artifacts. The removed secondary maxima and minima are shown in Figure 10 (marked by the arrows) for the case of a relatively normal EGG signal, which sometimes had a secondary peak within a cycle. The default value of 5.5 was used in this case.

This phase finishes the peak-picking procedure; the final maxima and minima defining the vocal fold oscillatory cycles (i.e., newlocmax5 and newlocmin5, Figure 10) are determined at this point.

The time indices of the finally selected maxima (newlocmax5) are used to derive the F_0 values for each cycle [4]. The F_0 values are further processed by the script *freqcheckJ.m*:

```
running function freqcheckJ.m
1) Lower frequency limit check (F0lower)
2) cycle-to-cycle variation check (change)
3) 3 consecutive non-zero values check
```

In this script, a condition is applied that all the F_0 values shall be higher than the lower F_0 limit specified at the beginning. If the F_0 value is lower than this limit, it is discarded by setting it to the value of zero. Next, a cycle-to-cycle variation check is applied to eliminate irregular peaks not related to vocalization. The F_0 of each three consecutive cycles are compared for this purpose ([4]; phase 12). Again, the F_0 value, which has not passed this condition, is discarded by setting its value to be zero. Finally, a condition is applied which says that a valid voicing segment shall contain at least three vibratory cycles. If there are less than 3 consecutive non-zero frequency values, they are discarded by setting their values to be zero.

The abovementioned operations provide the final maxima and minima, which define the oscillatory periods of the vocal folds and their fundamental frequency. The program displays a figure that shows the results of the F_0 and *SPL* analysis (Figure 11).

Next, the program asks whether to calculate the vocal doses (this happens only in the case when the calibration files, i.e., File 1 and 2 described earlier, were specified and the absolute *SPLs* were calculated):

Švec, Popolo & Titze: Goldilocks Passage and Scripts for Experimental Vocal Dose Measurement

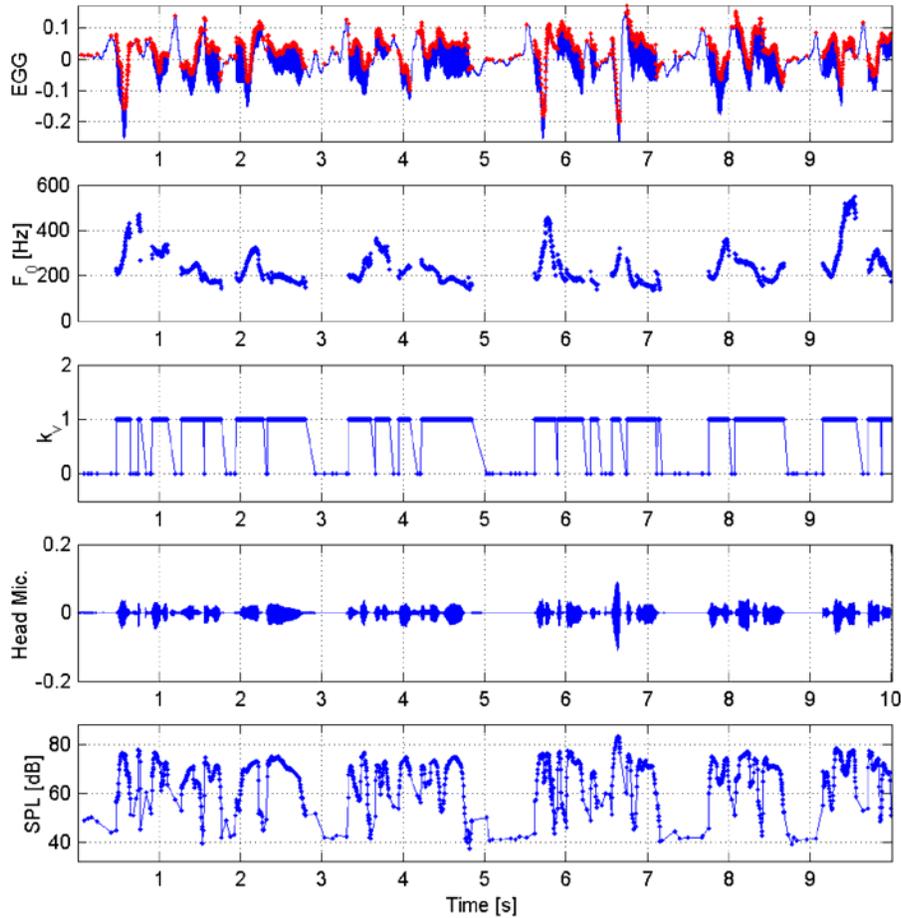


Figure 11. The results of the fundamental frequency (F_0) extraction and voicing detection (k_v) from EGG combined with the absolute SPL values obtained from the microphone signal. The F_0 and SPL values are obtained for each single extracted period of vocal fold vibration (for unvoiced segments with $k_v=0$, the F_0 values are not plotted). (File analyzed: *female_exag_reading.wav*).

Do you want to calculate doses using empirical rules?
 Enter 1 for YES or 0 for NO >

After entering “1 <Enter>”, the program wants to know, whether the measured subject was a male or a female (the empirical rules differ for male and female, see [4;6]):

Enter 1 for male subject or 2 for female subject:>

The National Center for Voice and Speech Online Technical Memo No.1, April 2003, version 1.4

After entering the relevant number (1 or 2), the program completes the computation and prints the computed vocal doses on the screen:

```
female_exag_reading_410.txt  
10-Apr-2003  3:51 PM
```

```
tmeas = 10  seconds (Total Measurement Time)  
voicing percentage = 56%
```

Dose calculations:

```
Dc = 1324 cycles (Cycle Dose)  
Dt = 5.61 seconds (Time Dose)  
Dd = 3.56 meters (Distance Dose)  
De = 1.354e+003 joules/m^3 (Dissipated Energy Dose)  
Dr = 2.613e-004 joules (Radiated Energy Dose)
```

The example above says that the analysis was done using a file *female_exag_reading.wav*, the results of the analysis were saved into a text file *female_exag_reading_410.txt* and the analysis was done on April 10, 2003. The total measurement time (i.e., the duration of the analyzed audio file) was 10 seconds, the voicing percentage was 56% and the Time Dose (i.e., the total time during which the vocal folds were vibrating) was 5.61 seconds. During the measurement time the vocal folds accomplished the total of 1324 oscillatory cycles (Cycle Dose) and traveled 3.56 meters on their oscillatory trajectory (Distance Dose). The total heat generated in the vocal folds due to viscous losses was $1.354 \cdot 10^3$ J/m³ (which corresponds to 1.354 mJ/cm³ of Dissipated Energy Dose) and the total energy radiated out of the mouth as a result of vocalization was 0.2613 mJ (Radiated Energy Dose).

Finally, the program displays two figures -- speech range profiles (Figure 12) and the cumulative doses for the particular analyzed file (Figure 13).

Svec, Popolo & Titze: Goldilocks Passage and Scripts for Experimental Vocal Dose Measurement

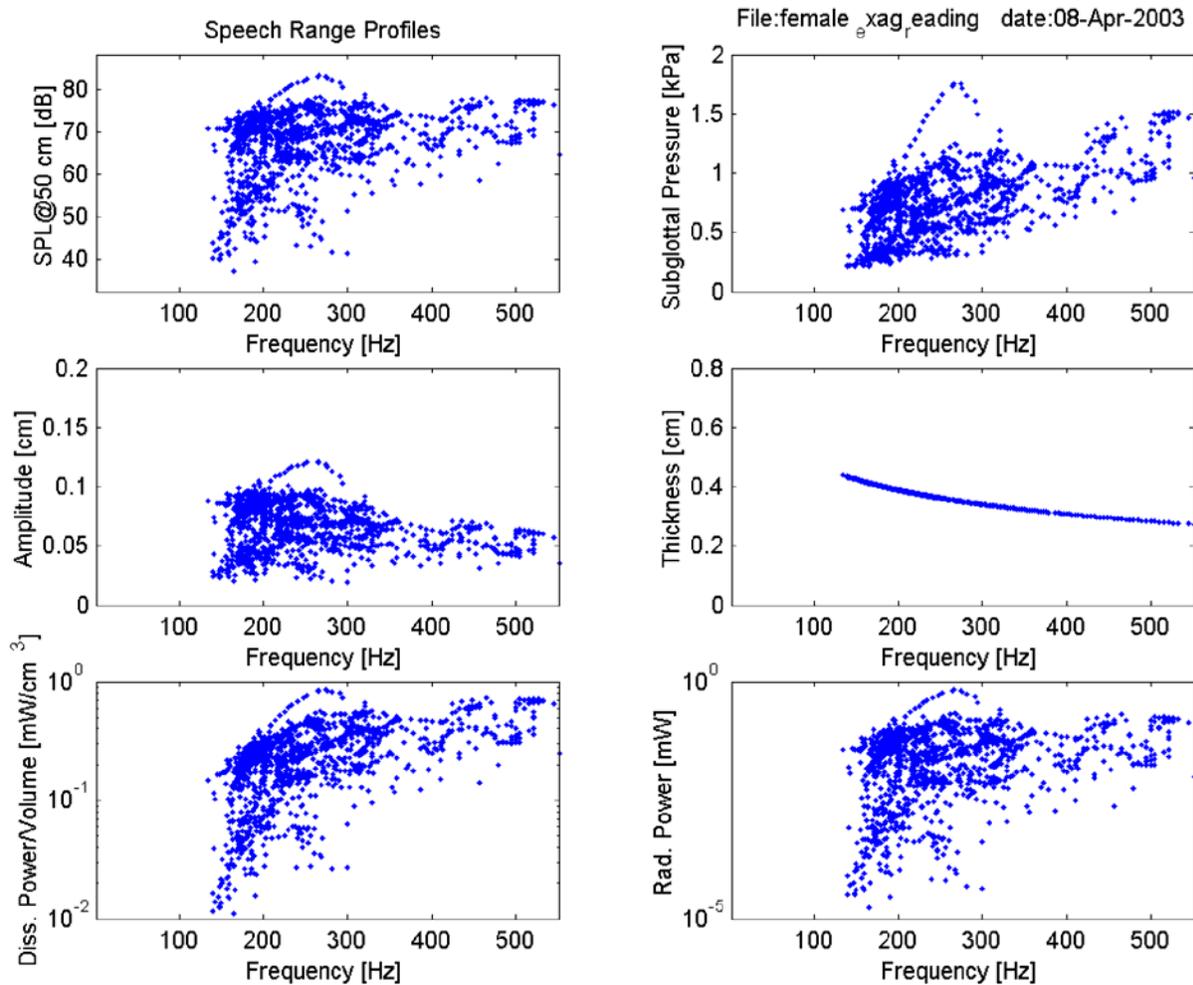


Figure 12: Speech range profiles displaying values derived for the individual periods of the vocal fold vibration. The following values are plotted against the fundamental frequency F_0 : absolute sound pressure level ($SPL@50\text{ cm}$), subglottal pressure, vocal fold amplitude, vertical thickness of the vocal folds, power dissipated in 1 cm volume of the vocal folds and power radiated from the mouth. (File analyzed: *female_exag_reading.wav*).

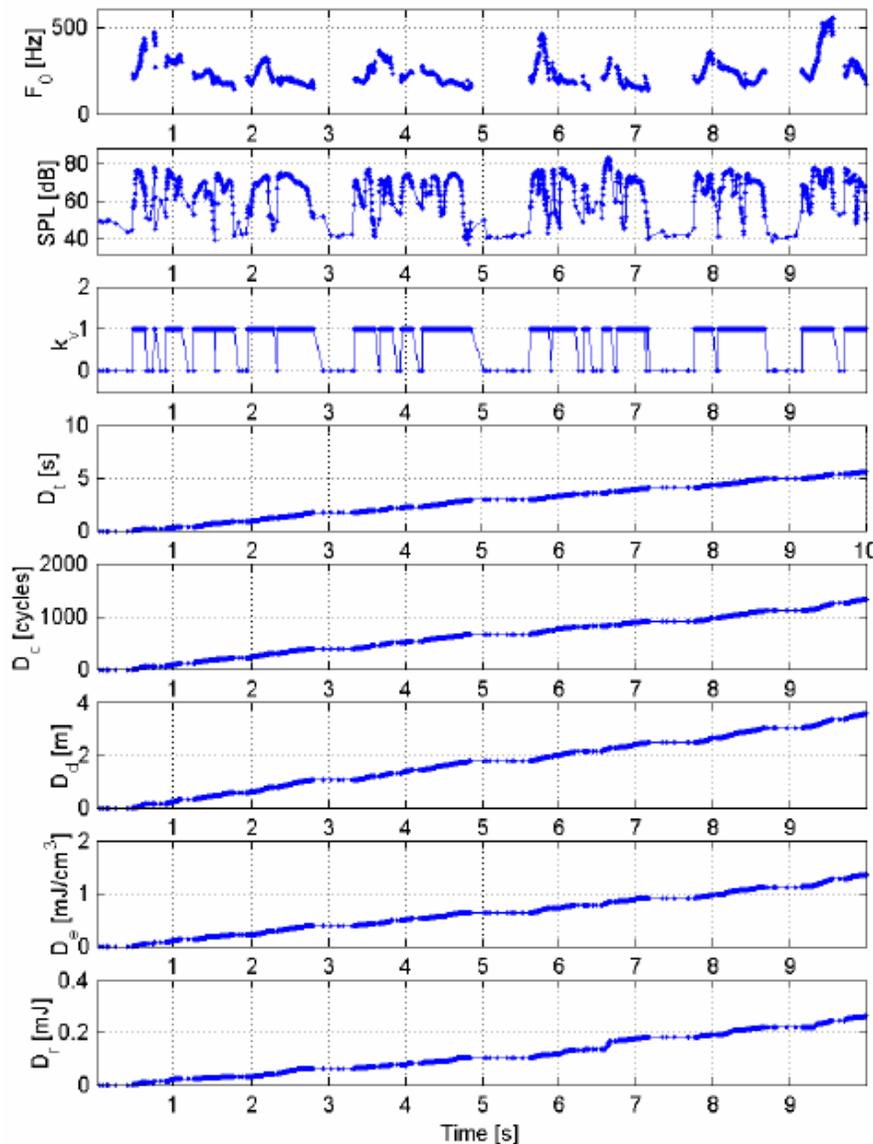


Figure 13. The cumulative vocal doses. The three upper graphs show the extracted values of F_0 , $SPL@50\text{ cm}$ and the voicing detection parameter k_v extracted from the speech signal. The graphs below show the five vocal doses (derived from the F_0 , SPL and k_v values) accumulating during speech. The final values of the doses after 10 seconds in the graph correspond to the final values printed by the script as given in the text: the *Time Dose* D_t reaches the value of 5.61 seconds after 10 seconds. The *Cycle Dose* D_c reveals that the vocal folds accomplished the total of 1324 cycles. Total distance of 3.56 meters traveled by the vocal folds on their oscillatory trajectory is measured by the *Distance Dose* D_d . The *Dissipated Energy Dose* D_e reaches the final value of 1.354 mJ/cm^3 and the total *Radiated Energy Dose* D_r is 0.2613 mJ. (File analyzed: *female_exag_reading.wav*).

Švec, Popolo & Titze: Goldilocks Passage and Scripts for Experimental Vocal Dose Measurement

Literature

- [1] ANSI S1.4-1983, American national standard specification for sound level meters Acoustical Society of America, 1985.
- [2] Eargle J. The microphone book. Boston: Focal Press, 2001.
- [3] Merhaut J. Theory of electroacoustics. New York/London: McGraw-Hill International Book, 1980. Translated from Czech by R. Gerber. (Czech version: Merhaut J. Teoretické základy elektroakustiky. 4 ed. Praha: Academia, 1985).
- [4] Švec JG, Popolo PS, Titze IR. Measurement of vocal doses in speech: Experimental procedure and signal processing. *Logopedics Phoniatrics Vocology*, 28(4): 181-192 (2003).
- [5] Titze IR. Workshop on acoustic voice analysis. Summary statement. National Center for Voice and Speech, 1995.
- [6] Titze IR, Švec JG, Popolo PS. Vocal dose measures: Quantifying accumulated vibration exposure in vocal fold tissues. *J Speech Lang Hear Res*. 46(4): 922-935 (2003).

Acknowledgements

The work has been supported by the National Institutes of Health, Project NIDCD 1, RO1 DC04224-01 (Research Towards Occupational Safety In Vocalization). 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: Albert Worley, Andrew Starr, John Nix, Kate Emerich and Jennie Kawulok.

Copyright

The MATLAB scripts, images and text enclosed in this memo and accompanying this memo are open to use by the public as a service of the National Center for Voice and Speech. However, we ask the reader to respect the time and effort put into this manuscript and research. If the text, images, or included scripts are used, the user agrees to reference to this document and the National Center for Voice and Speech. We reserve the right of refusal to (1) be authors on papers using the device and scripts and (2) have the supporting project acknowledged. The user agrees to freely share with the NCVS any extension software build on the scripts contained.

Revisions

- 1.4 Eric Hunter: minor edits, copyright (Apr. 2004)