

A Personal Voice Analyzer and Trainer

Markus Borgh^a, Sven Johansson^a, Åsa From^b, Fredric Lindström^c

^aBlekinge Institute of Technology, Department of Signal Processing, SE-37225, Ronneby, Sweden

^bLandstinget Blekinge, SE-37181, Karlskrona, Sweden

^cLimes Technology AB, Box 268, SE-90106, Umeå, Sweden

Abstract— This paper presents a personal voice analyzer and trainer that allow the user to perform four daily exercises to improve the voice capacity. The system grades how well the user is performing the exercises by analyzing the duration, the intensity and the pitch of the user's voice.

I. INTRODUCTION

Voice disorders such as dysarthria reduces the voice capacity. Dysarthria could be caused by neurological diseases such as Parkinson disease [1]. A way to improve the voice capacity is by performing daily exercises. This has been successful in treatment of Parkinson patients [2].

This paper presents four exercises that should be performed on a daily basis to improve the voice and speech performance. The exercises have been implemented in a portable consumer product, which measures and grades how well the user is performing the exercises. The grades range from 1 to 10, where a higher grade represents a better result.

II. EXERCISES

The exercises train different aspects of the voice. Exercise 1 and 4 focuses on duration and intensity while exercise 2 and 3 focuses on the pitch. This gives an all round training for the user to improve the voice. The exercises are based on common methods for voice therapy [2].

In exercise 1 the user is supposed to have a sustained phonation as long as possible. In exercise 2 the user should perform a glissando from a low frequency up to a high frequency. Exercise 3 also consists of a glissando but here the starting frequency is high and the exercise ends at a low frequency. In exercise 4 the user is supposed to produce a loud voice. A general criteria for all exercises is that the user is required to produce a sufficiently loud sound to activate the measurement.

III. GRADE CALCULATION

The user voice is picked up by a microphone generating the microphone signal $m(n)$, where n is the sample index. The average power $P(k)$ in a block of size N is

$$P(k) = \frac{1}{N} \sum_{i=0}^{N-1} m^2((k+1)N - i), \quad (1)$$

where k is the block index. The active voice $F(k)$ is defined as

$$F(k) = \begin{cases} 1 & \text{if } P(k) > T_p \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where T_p is the power threshold that the user has to surpass to start the measurement.

A. Exercise 1

The grade, G_1 , is proportional to how long time the user is able to produce a sound above a certain threshold, i.e.

$$G_1 \propto I_{1,e} - I_{1,s}, \quad (3)$$

where the measurement interval is decided by the start time

$$I_{1,s} = \min \left(k \in [0, \infty) : \sum_{i=0}^{M-1} F(k+i) = M \right) \quad (4)$$

and the end time

$$I_{1,e} = \min \left(k \in (I_{1,s}, \infty) : \sum_{i=0}^{M-1} F(k+i) = 0 \right), \quad (5)$$

where M is a constant determining the time the user is required to be active/nonactive before the grade calculations begins/ends.

B. Exercises 2 and 3

The grades, G_2 and G_3 , in exercises 2 and 3 are proportional to the mean square error between the measured fundamental frequency, $f_0(k)$, and an optimal (straight line) frequency $f_l(k)$. The fundamental frequency $f_0(k)$ is measured with a pitch extracting algorithm [3]. In this exercise, $f_0(k)$ and $f_l(k)$ ranges from a preset constant lower boundary f_{low} to a preset upper boundary f_{high} . The grade is

$$G_x \propto 10 - \frac{e_{x,\text{mean}}}{I_{x,e} - I_{x,s}}, \quad (6)$$

where x is 2 or 3 and the mean square error $e_{x,\text{mean}}$ is defined as

$$e_{x,\text{mean}} = \sum_{i=I_{x,s}}^{I_{x,e}} [f_0(i) - f_l(i)]^2, \quad (7)$$

where the optimal frequency is

$$f_l(k) = k \frac{f_{high} - f_{low}}{|I_{x,e} - I_{x,s}|} + f_{low}. \quad (8)$$

The measurement interval in exercise 2 is determined by the start time

$$I_{2,s} = \min \left(k \in [0, \infty) : \sum_{i=k}^{k+T_2} F(i) = T_2 \text{ and } f_{\text{low}} - C_2 \leq F(k)f_0(k) \leq f_{\text{low}} \right), \quad (9)$$

where C_2 is a constant and T_2 is the required phonation time, and the end time

$$I_{2,e} = \min \left(k \in (I_{2,s} + T_2, \infty) : F(k)f_0(k) \geq f_{\text{high}} \right). \quad (10)$$

The measurement interval in *exercise 3* is determined by the start time

$$I_{3,s} = \min \left(k \in [0, \infty) : \sum_{i=k}^{k+T_3} F(i) = T_3 \text{ and } f_{\text{high}} \leq F(k)f_0(k) \leq f_{\text{high}} + C_3 \right), \quad (11)$$

where C_3 is a constant and T_3 is the required phonation time, and the end time

$$I_{3,e} = \min \left(k \in (I_{3,s} + T_3, \infty) : F(k)f_0(k) \leq f_{\text{low}} \right). \quad (12)$$

C. Exercise 4

The grade, G_4 , in this exercise is proportional to the average power P_{ave} , i.e.

$$G_4 \propto P_{\text{ave}}, \quad (13)$$

where the average power is given by

$$P_{\text{ave}} = \frac{\sum_{i=I_{4,s}}^{I_{4,e}} F(i)P(i)}{\sum_{i=I_{4,s}}^{I_{4,e}} F(i)}, \quad (14)$$

where the summation interval, $[I_{4,s}, I_{4,e}]$, is the longest interval where the active voice $F(k)$ is 1, i.e.

$$I_{4,s} = \min \left(k \in [0, \infty) : \sum_{i=k}^{k+T_4} F(i) = T_4 \right) \quad (15)$$

and

$$I_{4,e} = \max \left(k \in (I_{4,s} + T_4, \infty) : \sum_{i=I_{4,s}}^{k-1} F(i) = k - I_{4,s} \right), \quad (16)$$

where T_4 is the required phonation time.

IV. SUBJECTIVE EVALUATION

The voice analyzer have been tested by a test group in both a laboratory environment and in a home environment. In the laboratory environment some parameter values were determined that gave a good performance. These parameter values are listed in table I. It was also concluded that the frequency values f_{low} and f_{high} should be set individually for each person. The home environment study indicated that the

TABLE I
PARAMETER VALUES ESTABLISHED DURING THE EVALUATION PERIOD.

Parameter	Value
T_p	65 dB
M	1.3 s
T_2	2 s
T_3	2 s
T_4	4.1 s
C_2	5 Hz
C_3	0 Hz

exercises were easier to perform thanks to the feedback that the system delivered to the user. Another positive effect is that the user is encouraged to perform the daily exercises by the desire to receive a higher grade.

V. CONCLUSION

This paper presents four exercises, implemented in a consumer product, designed for voice training. To each exercise a grade is calculated that shows how well the user performed the exercise. The product has not been large scale tested yet but the result from a test group indicates that the exercises are easier and more fun to perform thanks to the grading system.

REFERENCES

- [1] J. Logemann, H. Fisher, B. Boshes, and E. Blonsky, "Frequency and concurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients," *Journal of Speech and Hearing Disorders*, vol. 42, pp. 47–57, 1978.
- [2] L. Raming, A. Pawlas, and S. Countryman, *The Lee Silverman Voice Treatment (LSVT®): A practical guide to treating the voice and speech disorders in Parkinson disease*, Iowa City, LSVT Foundation, 1995.
- [3] U Zölzer, *DAFX: Digital Audio Effects*, pp. 337–349, Wiley, 2002.