An Easy Approach for the Classification of Children's Voices based on the Fundamental Frequency Estimation

Laura Verde, Giuseppe De Pietro and Giovanna Sannino

Institute of High Performance Computing and Networking (ICAR) - CNR, 80131 Naples, Italy

Keywords: Dysphonia, Children Voice Disorders, m-health Application, Fundamental Frequency Estimation.

Abstract: Voice disorders, also called dysphonia, are qualitative and quantitative alterations of the voice. These pathologies, unfortunately, affect from 6% to 38% of children in the world. Voice disorders may have a negative impact on communication effectiveness, social development and self-esteem. The first weapon against the diffusion and the worsening of these pathologies is prevention. Acoustic analysis is one of the most important tools to appraise the state of health of a voice. It provides information about the possible presence of voice disorders by evaluating specific parameters like the Fundamental Frequency. In this paper we present an easy approach based on a mobile application for voice screening in children. The app provides a robust methodology for the fundamental frequency estimation of the voice signal by analysing in real time a child's signal. It consists of a continuous vocalization of the vowel /a/ of five seconds in length. The methodology is also able to evaluate undesired noise that can alter the Fundamental Frequency estimation and the correct classification of the evaluated voice signal as pathological or healthy.

1 INTRODUCTION

Dysphonia indicates a disturbance of the phonatory apparatus, that can alter vocal quality, pitch, and loudness. Such disorders can limit the conversations between people (Glaze, 1996) and social relationships and reduce self-esteem.

Although people think that these disorders mainly involve adults, particularly specific groups of professional voice users such as teachers or singers, dysphonia also affects children. In fact, reports of childhood voice disorders describe an incidence among school-age children (5-18 years), with a percentage between 6% and 38%. Kahane and Mayo (Kahane and Mayo, 1989) concluded that very few children with voice disorders, only between 2% and 4%, are ever seen by a Speech-Language Pathologist (SLP) (Deal et al., 1976; McNamara and Perry, 1994). Leeper (Leeper, 1992) estimates that 38% of elementary school children present with chronic hoarseness. There is a wide variety in the numbers relating to this incidence due to: 1) a lack of consistent measurement techniques, and 2) a variability in listener perceptual judgement.

In children dysphonia may have a multifactorial origin: various causes, in fact, can be combined with each other and contribute to the phenomenon. Voice disorders can have an organic nature, resulting from, for example, chromosomal defects, congenital anomalies, lesions of the larynx, inflammatory reactions of the laryngeal mucosae and deep tissues, endocrine disorders (errors of the metabolism whose incidence on normal enzymatic sequences can cause an abnormal infiltration or faulty nerve and muscle function) or gastroesophageal reflux (Dejonckere, 1999). Generally, voice diseases in children can be associated with vocal abuse. Typical behaviours of children such as talking too long, too loud and with too much effort, can cause damage to the vocal folds. Often, these behaviours are influenced by the environments where the children spend significant time, such as dry or noisy rooms in schools, or by the environmental background noise of loud television or music in their rooms.

In detail, paediatric voice disorders can be classified into two groups:

- **Congenital**: Vocal Fold Paralysis, Laryngeal Stenosis, Laryngomalacia, Laryngocele, Webbing, and Anterior Laryngeal Cleft;
- **Acquired**: Chronic Laryngitis, Laryngeal Trauma, Hyperfunction w/o Lesions, Vocal Nodules, Vocal Polyps, Contact Ulcers, or Vocal Fold Paralysis.

There is no doubt that prevention is the most im-

Verde, L., Pietro, G. and Sannino, G.
An Easy Approach for the Classification of Children's Voices based on the Fundamental Frequency Estimation.
DOI: 10.5220/0005849005700577
In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2016) - Volume 5: HEALTHINF*, pages 570-577
ISBN: 978-989-758-170-0

portant weapon for the future. The increase of information about these disorders, the abolition of negative environmental factors and of unhealthy behavioural habits, and the use of tools for voice analysis are the main factors to avoid the contracting of any of these pathologies and represent the prerequisite for a successful application of the therapy.

In this paper we propose a methodology able to estimate the presence of voice disorders in a non-invasive way in order to provide an easy, fast and entertaining tool, using a mobile health application. This instrument can be useful to monitor the status and progress throughout the treatment program, as well as to provide useful advice on healthy lifestyle behaviours to follow to prevent voice disorders.

2 BACKGROUND

The SIFEL protocol (Lucchini, 2002) indicates the guidelines for evaluating the presence of voice disorders. The Italian Society of Logopedics and Phoniatrics developed this protocol in accordance with the directives of the Committee for Phoniatrics of the European Society of Laryngology.

According to this protocol, the evaluation of dysphonia consists in a series of tests, such as an anamnestic evaluation, a laryngo-video-stroboscopic examination, a subjective self-assessment of the voice and an acoustic analysis. A laryngo-video-stroboscopic examination is necessary to identify morphodynamic changes of the larynx. It is difficult to perform this invasive examination on children due to the complaints and inconveniences that they may cause. For this reason, acoustic analysis is the most effective instrument to extract of the pathological voice characteristics.

This analysis provides a view of some characteristics of the speech signal, by calculating some parameters like the Fundamental Frequency (F_0), jitter or shimmer, parameters that can be quantified at that specific time and whose evolution can be monitored over time. In particular, the F_0 provides a measure of the rate of vocal vibration, any lesions on the vocal folds being able to alter the F_0 value (Casper and Leonard, 2006).

A voice disorder is generally present if the F_0 value, as well as the other parameters calculated in the analysis acoustic, is outside an appropriate healthy range. Unfortunately, the F_0 is influenced by different factors:

- anatomic factors: child, female and male voices differ significantly due to changes in the anatomical structures of the larynx and especially of the vocal folds during the years (Angelillo et al., 2015; Hunter et al., 2011);
- the physical conformation of the user: the length, tension, or mass of the vocal folds (Hunter et al., 2011);
- the use of the voice: the pressure of the forced expiration, or the sub-glottal pressure (Hunter et al., 2011);
- the state of health of the person: the emotions and state of health of people can determine changes in the vibrations of vocal folds and so alter the dynamics of pitch (Johnstone and Scherer, 1999; Nerrière et al., 2009);
- lifestyle: incorrect lifestyle habits like smoking (Gonzalez and Carpi, 2004), and alcohol intake (Cooney, 1998), such bad habits unfortunately increasing among boys and girls (Lorant et al., 2015; Pinilla et al., 2002; Simons-Morton et al., 2001) and so influencing the health of the voice.

Therefore, it is important to perform the F_0 estimation as accurately as possible to obtain a reliable acoustic analysis. The mobile application realized provides the F_0 estimation, the first and the most important parameter of the acoustic analysis.

We have aimed to develop a gamification instrument useful to perform an evaluation of the possible presence of voice disorders, using a simple mobile device such as a smartphone or tablet. In this way, children can independently assess the health of their own voice. Several studies (King et al., 2013; Miller et al., 2014) have reported, in fact, that mobile health applications can represent an effective way to promote health interventions. The choice of using mobiles has been dictated by the rapid spread of these devices among boys and girls and the continued development of m-Health systems. Given children's propensity for mobile apps, this approach may provide important opportunities to engage them and to help them follow healthy lifestyles and adopt appropriate behaviours.

3 RELATED WORK

Several systems and apps have been found in literature, developed to promote the prevention of health disorders and to educate towards a correct lifestyle. There are commercial games, for example, aimed at increasing physical fitness or at counteracting depression in teenagers and social isolation in the elderly (McCallum, 2012). Games have also been realized for specific health conditions, such as Bant, a mobile app useful to improve glucose monitoring among adolescents with diabets (Cafazzo et al., 2012).

There are, moreover, several apps to help people suffering from dementia, as reported by (Kong, 2015). In (Kong, 2015)'s study the performance of several appropriate apps, present on the iTunes market, has been evaluated with people suffering from early stage dementia. Different parameters were analyzed such as the usability, price of the app, and the reactions of clinicians and involved patients.

In this study, for the F_0 estimation, several systems and apps, based on different algorithms found in literature, have been studied. Nevertheless, none of these can be considered as a personal and portable instrument that is reliable in terms of its results and pathology classification capability.

Opera Vox (Baki et al., 2013) is an example of an app allowing people to perform acoustic measurements. Unfortunately, it indicates to the user only the values of the calculated parameters of the acoustic analysis, these results not being easily interpretable by users without the support of an expert. In Opera Vox, the used algorithm to estimate the F_0 is based on the autocorrelation function (ACF).

The ACF (Tan and Karnjanadecha, 2003) of the speech signal is a traditional algorithm for the F_0 estimation. It is used to estimate the fundamental period, selecting the maximum peak of this function. The ACF is obtained by computing the correlation between a part of the windowed signal and its shifted segment.

It is used also by PRAAT (Boersma, 1993) to estimate the F_0. PRAAT is a tool, distributed for free use, commonly employed for acoustic analysis in clinical and research settings. In the PRAAT F_0 estimation algorithm, the speech signal is divided into frames using an appropriate window to minimize spectral leakage. The F_0 is estimated for each frame.

An alternative to the autocorrelation function is the "Average Magnitude Difference function (AMDF)" (Ross et al., 1974), where the fundamental period is estimated by calculating the local minima of the difference function between the speech signal and its shifted version of an appropriate period.

Another algorithm is the "Dynamic Programming Projected Phase-Slope Algorithm (DYPSA)", able to perform an automatic estimation of glottal closure instants (GCIs) in voiced speech. It employs dynamic programming to identify the best GCI candidates by minimizing some cost functions (Naylor et al., 2007).

The "Robust Algorithm for Pitch Tracking (RAPT)" is a time-domain F_0 estimation algorithm that uses the "Normalized Cross-Correlation Function (NCCF)" (Talkin, 1995). It compares frames of the original speech with sub-sampled frames of the same signal and searches for the local maxima of the NCCF to identify the peak locations and amplitude estimates. On all frames of the speech signal the series of NCCF peaks and the F_0 candidates are selected by using a dynamic programming. While RAPT estimates F_0 in the time domain, an algorithm that works in the frequency domain is SHRP (Sun, 2002), in which the F_0 estimate is obtained through spectrum shifting on a logarithmic frequency scale, calculating the Subharmonic-to-Harmonic Ratio (SHR).

The "Sawtooth Waveform Inspired Pitch Estimator (SWIPE)" (Camacho and Harris, 2008), instead, is an algorithm in which the fundamental frequency of the sawtooth waveform whose best spectrum equals the spectrum of the speech signal is adopted as the pitch.

Even if there are many suitable algorithms for the fundamental frequency estimation, most of these are not runnable on mobile devices and are not easy to use as prevention tools by non-experts.

4 THE MOBILE APPLICATION

The mobile application was developed in the Java Programming Language using Eclipse IDE and the Android Software Developer Kit (SDK). Several aspects were considered in the realization of the app. We aimed to develop an app that boys and girls can use easily, with a clear design and an intuitive interface to perform several tasks.

The main objectives, on which the mobile application was developed, are:

- to increase the user's motivation for self-improvement;
- to educate the user in the physical behaviours that contribute to an inappropriate voice (e.g. posture, breathing, and muscular tension);
- to inform the user about lifestyle factors that contribute to an inappropriate voice (e.g. a noisy environment, sleeping or eating habits, and air pollution);
- to inform the user about interpersonal behaviours that contribute to an inappropriate voice (e.g. talking too much, ignoring feedback, and competing for attention);
- to estimate and evaluate the fundamental frequency of a voice signal;
- to discriminate a possible pathological voice from a healthy one.

To achieve the latter two objectives, we optimized a methodology for F_0 estimation, with a noise evalua-

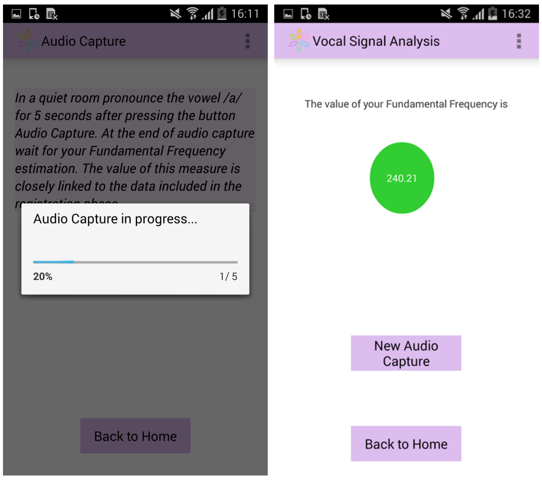

Figure 1: Screenshots of the audio capture phase and of the vocal signal analysis report.

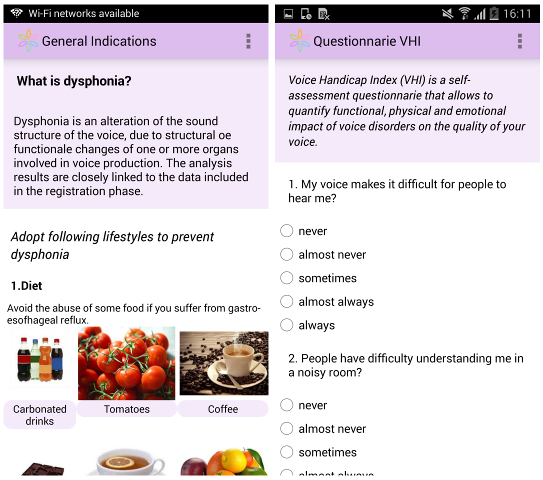

Figure 2: Screenshots of the screen containing information about voice disorders, e.g. about dysphonia and of the VHI Questionnarie.

tion, using an appropriate methodology, described in the following sections.

The realized app acquires the child's vocalization. It consists of recording of the vowel /a/ of five seconds in length as required by the SIFEL protocol. This speech signal is elaborated to estimate the F_0 by reducing the noise in real time, and to classify the child's voice as possibly pathological or healthy. The procedures of audio capture and F_0 estimation are reported in the screenshots of Figure 1.

Moreover, the realized app provides information about dysphonia to explain to boys and girls its causes and to suggest are suitable preventative healthy life-style behaviours, e.g. avoiding certain kinds of food, shouting, temperature changes, or noisy environments. Figure 2 reports the screen containing information about dysphonia and some healthy life-style behaviours.

Finally, the app allows the user to complete two questionnaires useful for the self-evaluation of the presence of a voice disorder, as required by the SIFEL protocol: the Voice Handicap Index (VHI) (Forti et al., 2014) and the Reflux Symptom Index (RSI) (Belafsky et al., 2002). The first questionnaire is a correct and complete instrument to evaluate the patient's self-perception of his her voice disorders. The second questionnaire, instead, estimates if the patient suffers from extra-esophageal reflux, a risk factor for dysphonia. An example of the VHI questionnarie is provided in the screenshot of Figure 2, each question in the questionnaire including five possible answers to to which a specific score corresponds depending on the severity of the indicated symptom.

The user, to access the functionalities of the app and perform the described procedures, must be authenticated by the system, thanks the insertion of his/her credentials in the Login phase, shown in the screenshot of the Figure 3. These credentials are recorded and saved in the Registration phase with the Wellness Server, an operation that the user must perform at his/her first access to the app. The connection with the Wellness Server is useful to collect the monitored data of the user, to improve the his/her well-being. The user, after being authenticated, can access the homepage of the app where all the functionalities are shown. A screenshot of the homepage of the app is shown in Figure 3.

4.1 The Fundamental Frequency Estimation Algorithm

The developed methodology for the F_0 estimation is based on the algorithm reported in (De Cheveigné and Kawahara, 2002). By hypothesizing the periodicity of the speech signal x_t with a period T in each time-segment of the signal, called the window (W), that in our methodology is 10 ms long, we can define that the speech signal does not vary for a time shift of T.

To find the unknown fundamental period T of the speech signal, τ values that minimize the difference function $d_t(\tau)$ are searched. The sum of the squared differences between the speech signal and its shifted version, in every window in which the speech signal is divided, identifies the difference function $d_t(\tau)$, that is:

$$d_t(\tau) = \sum_{j=1}^{W} (x_j - x_{j+\tau})^2 \tag{1}$$

Due to the non perfect periodicity of the speech signal, caused by to physiological and intentional problems, this function is sensitive to amplitude

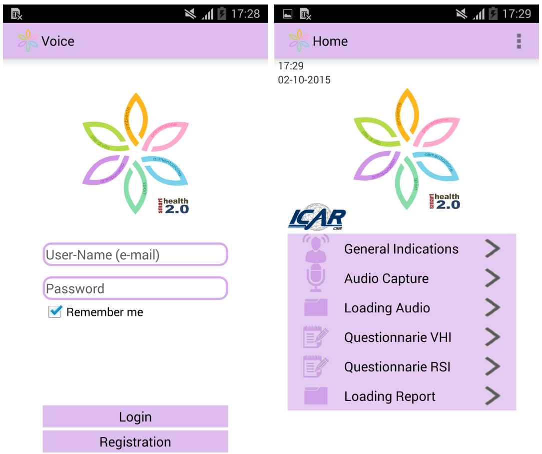

Figure 3: Screenshots of the log-in and of Homepage.

changes of the signal. A cumulative difference function is applied to reduce this effect and improve the F_0 estimation, defined as follows:

$$d_t'(\tau) = \begin{cases} 1 & \text{if } \tau = 0 \\ \frac{d_t(\tau)}{\frac{1}{\tau}\sum_{j=1}^{\tau} d_t(j)} & \text{else} \end{cases} \tag{2}$$

Therefore, the values of τ that minimize this cumulative difference function are searched and among these those smaller than a threshold value are considered to calculate the unknown period. In our proposed methodology, this threshold value was found empirically, a series of experimental tests have been carried out to find the value that gives the best estimation of F_0, also when the noise is inadvertently added to the useful signal for the acoustic analysis. The threshold value reported in (De Cheveigné and Kawahara, 2002) is equal to 0.10, while the conducted experimental tests have shown that the value that provides the best estimation of the F_0 is equal to 0.40. To increase the accuracy of the F_0 estimation, a parabolic interpolation was used to refine the local minima of $d_t'(\tau)$ with their neighbouring values.

The F_0 is calculated as the average of the fundamental frequencies of all the windows into which the speech signal has been divided, obtained as the inverse of T_i on all the windows, that is:

$$F_0 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_i} \tag{3}$$

indicating N as the number of windows.

4.2 The Noise Evaluation

The F_0 estimation can be altered by the introduction of noise during the speech signal acquisition. Therefore, the acoustic analysis can be affected by possible errors due to increase in the potential number of false-positive diagnoses of voice disorders. Consequently, we aimed to reduce the noise effects added during the child's vocal signal acquisition.

This objective was achieved using an FIR filter. We have developed a causal linear-phase FIR filter using the windowing technique, multiplying an ideal filter with a finite-duration window function.

In practice, several types of windows are commonly used. To realize our filter we have used the Hanning window, but there are other windows available, such as rectangular, Hamming or Blackmann.

The following equation describes the realized Hanning filter in the time domain:

$$y(n) = d(x(n) + 2x(n-1) + x(n-2)) \tag{4}$$

where the output signal is indicated as $y(n)$ and the input signal as $x(n)$, while d is the normalization factor. In this work, we have found the value of d empirically. It is equal to 2 and it was discovered thanks to experimental tests to identify the value that gives the best noise evaluation.

5 EXPERIMENTS

The performance of the methodology for the F_0 estimation was evaluated in a testing phase. This performance was compared with other existing algorithms, Praat (Boersma, 1993), the tool described in (Ross et al., 1974), SWIPE (Camacho and Harris, 2008) and Yin algorithms (De Cheveigné and Kawahara, 2002).

The results are reported in terms of sensitivity, specificity and accuracy, on the basis of these definitions:

- True Positive (TP): the algorithm recognized the pathology when the speech signal was pathological;
- True Negative (TN): the algorithm recognized the speech signal as healthy when it was healthy;
- False Positive (FP): the algorithm recognized the pathology when the speech signal was healthy;
- False Negative (FN): the algorithm recognized the speech signal as healthy when it was pathological.

In detail, the *sensitivity* measures the proportion of true positive users (pathological voices) that are correctly identified as positives by the evaluation of the proposed methodology. Its value is evaluated as TP/(TP+FN). The *specificity*, instead, is defined as the ratio of true negative users (healthy voices), users that are correctly identified as negative, to the total unaffected patients tested. This value is equal to

TN/(TN+FP). Finally, the *accuracy* is the proportion of the total number of correct predictions and it is equal to (TP+TN)/(TP+TN+FP+FN).

The tests were performed on voices from an available on-line database, the "Saarbrucken Voice Database" (SVD) (Martínez et al., 2012), downloaded from the URL [http://www.stimmdatenbank.coli.uni-saarland.de].

The SVD database is a collection of recordings of /a/ vowels. These recordings were made by the Institute of Phonetics of Saarland University. The fidelity of the signal was preserved by recording in a mono-channel in the WAVE format and sampling the recordings at 50 Hz with a resolution equal to 16-bit.

5.1 The Dataset

We have built a dataset composed of all the 22 children's voice samples of the SVD database, all the available voices in the database used. It contains the sustained phonation of the vowel sound /a/, the phonation used in the acoustic analysis as required by the SIFEL protocol. In detail the built dataset includes:

- 6 healthy voices, 5 voices of healthy male children and 1 voice of a healthy female child;
- 16 pathological voices, 6 voices of pathological male children and 10 of pathological female children.

The selected children are aged between 9 and 16 years.

In particular, the pathologies included: psychogenic aphonia, an inability to produce the voice, laryngitis, an acute or chronic inflammation of the vocal folds, and rhinolalia, an alteration of the voice timbre, which acquires a nasal character.

5.2 Results

The selected voices are classified as healthy or pathological considering a healthy range of values, the range being reported in (Nicollas et al., 2008). In particular, we used the healthy range from 235 to 270 Hz for male children and the range from 240 to 260 Hz for female children.

Voices samples that fall within these ranges are considered healthy. Those outside are considered as possibly pathological.

The performance of the developed methodology is reported in Table 1 in terms of sensitivity, specificity and accuracy, in comparison with the performance of other algorithms exiting in literature. The performance of all algorithms have been evaluated on the same dataset, composed by the selected voices from the SVD database, as indicated in the subsection 5.1.

Table 1: Results.

Algorithm	**Sensitivity** (%)	**Specificity** (%)	**Accuracy** (%)
Proposed Methodology	68.75	16.66	54.54
Praat (Boersma, 1993)	50.00	16.66	40.90
AMDF (Ross et al., 1974)	43.75	16.66	36.36
SWIPE (Camacho and Harris, 2008)	56.25	16.66	45.45
YIN (De Cheveigné and Kawahara, 2002)	50.00	16.66	40.90

The results reported in Table 1 indicate the good accuracy of our methodology in discriminating healthy voices from pathological ones compared to Praat, to AMDF-based tool, to SWIPE and Yin algorithm performances. Moreover, the table shows the high sensitivity (the average sensitivity value is equal to about 69%) of our methodology in comparison with the other algorithms. This means that the number of false negatives is lower, the algorithm generally recognizing the presence of a pathology when the speech signal is indeed pathological.

It is important to note that this methodology was embedded in a mobile application, usable and interpretable by people without any medical support, while the other algorithms are runnable only on Matlab or are embedded in proprietary desktop applications, able to provide numerical results of the F_0 that can be interpreted only by medical experts.

Although the proposed methodology performs an accurate classification of voice disorders in comparison with other algorithms, it does not consider different factors that can alter voice production, such as the anatomical conformation, state of health and lyfestyle habits (smoking or alcohol intake). For this reason to improve the classification of voice disorders further studies will be addressed at estimating other parameters like jitter and shimmer, in accordance with the SIFEL protocol.

6 CONCLUSIONS

The diffusion of voice disorders at the paediatric age has been increasing over the last few years. However, fortunately, in contrast to past practice, a greater importance has been given to voice screening in children.

In most cases, voice disorders may impact on a child's state of health, and social and educational development. Therefore, it is important to diagnose of dysphonia early, without underestimating its symptoms and causes. In practice, many young people turn to a speech specialist only belatedly to resolve the pathology.

For this reason, in this paper we have presented an easy approach based on a mobile application for voice screening in children. The app provides a robust methodology for the fundamental frequency estimation of the speech signal on the recording of the vowel /a/ of five seconds in length, as provided by the protocol, classifying a voice as healthy or pathological. The methodology is also able to evaluate undesired noise that can introduce errors in the F_0 estimation, altering the classification of state of the vocal health.

The results obtained with the proposed methodology have been compared with the performance of other algorithms exiting in literature, Praat, a software used in clinical practice, an AMDF-based tool, SWIPE and Yin. The results of the testing phase have demonstrated that the distinction between healthy voices and pathological ones is performed with a good accuracy using the proposed methodology.

The developed app does not provide a diagnosis, our aim being provide an instrument for a first screening test, an easy and gamified instrument that can be used by children, suggesting a consultation with a qualified speech therapist for an appropriate diagnosis.

As our future plans, we would like to investigate gamification techniques to motivate children in the use of the mobile app. In detail, we aim to develop a game-based educational app to facilitate the learning phase with children, for example to distinguish between healthy and unhealthy foods. Moreover, gamification techniques will also be adopted to encourage children to complete the signal acquisition and, in the case of a prescribed therapy, to improve the motivation of children to practise home-based exercises.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the project "Smart Health 2.0" PON04A2_C for their support of this work. Additionally, the authors wish to thank Prof. Pierangelo Veltri, University "Magna Graecia" of Catanzaro (Italy), and Prof.Nicola Lombardo, Department of Otolaryngology-Head and Neck Surgery of the University "Magna Graecia" of Catanzaro (Italy) involved in the SmartHealth 2.0 project, for his useful contribution to the identification of the healthy range of values of F0 used in this study.

REFERENCES

Angelillo, I. F., Di Costanzo, B., Costa, G., Barillari, M., and Barillari, U. (2015). Epidemiological study on vocal disorders in paediatric age. *Journal of preventive medicine and hygiene*, 49(1).

Baki, M. M., Wood, G., Alston, M., Ratcliffe, P., Sandhu, G., Rubin, J. S., and Birchall, M. A. (2013). Comparison between operavox and mdvp: Preliminary results. *Otolaryngology–Head and Neck Surgery*, 149(2 suppl):P203–P204.

Belafsky, P. C., Postma, G. N., and Koufman, J. A. (2002). Validity and reliability of the reflux symptom index (rsi). *Journal of Voice*, 16(2):274–277.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the institute of phonetic sciences*, volume 17, pages 97–110. Amsterdam.

Cafazzo, J. A., Casselman, M., Katzman, D. K., and Palmert, M. R. (2012). 133. bant: An mhealth app for adolescent type i diabetes–a pilot study. *Journal of Adolescent Health*, 50(2):S77–S78.

Camacho, A. and Harris, J. G. (2008). A sawtooth waveform inspired pitch estimator for speech and music. *The Journal of the Acoustical Society of America*, 124(3):1638–1652.

Casper, J. K. and Leonard, R. (2006). *Understanding voice problems: A physiological perspective for diagnosis and treatment*. Lippincott Williams & Wilkins.

Cooney, O. (1998). *Acoustic analysis of the effects of alcohol on the human voice*. PhD thesis, Dublin City University.

De Cheveigné, A. and Kawahara, H. (2002). Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930.

Deal, R. E., McClain, B., and Sudderth, J. F. (1976). Identification, evaluation, therapy, and follow-up for children with vocal nodules in a public school setting. *Journal of speech and hearing disorders*, 41(3):390–397.

Dejonckere, P. (1999). Voice problems in children: pathogenesis and diagnosis. *International journal of pediatric otorhinolaryngology*, 49:S311–S314.

Forti, S., Amico, M., Zambarbieri, A., Ciabatta, A., Assi, C., Pignataro, L., and Cantarella, G. (2014). Validation of the italian voice handicap index-10. *Journal of Voice*, 28(2):263–e17.

Glaze, L. E. (1996). Treatment of voice hyperfunction in the pre-adolescent. *Language, Speech, and Hearing services in schools*, 27(3):244–250.

Gonzalez, J. and Carpi, A. (2004). Early effects of smoking on the voice: A multidimensional study. *Medical Science Monitor*, 10(12):CR649–CR656.

Hunter, E. J., Tanner, K., and Smith, M. E. (2011). Gender differences affecting vocal health of women in vocally demanding careers. *Logopedics Phoniatrics Vocology*, 36(3):128–136.

Johnstone, T. and Scherer, K. R. (1999). The effects of emotions on voice quality. In *Proceedings of the XIVth International Congress of Phonetic Sciences*, pages 2029–2032. University of California, Berkeley San Francisco.

Kahane, J. C. and Mayo, R. (1989). The need for aggressive pursuit of healthy childhood voices. *Language, Speech, and Hearing Services in Schools*, 20(1):102–107.

King, D., Greaves, F., Exeter, C., and Darzi, A. (2013). gamification: Influencing health behaviours with games. *Journal of the Royal Society of Medicine*, 106(3):76–78.

Kong, A. P.-H. (2015). Conducting cognitive exercises for early dementia with the use of apps on ipads. *Communication Disorders Quarterly*, 36(2):102–106.

Leeper, L. H. (1992). Diagnostic examination of children with voice disordersa low-cost solution. *Language, Speech, and Hearing Services in Schools*, 23(4):353–360.

Lorant, V., Soto, V. E., Alves, J., Federico, B., Kinnunen, J., Kuipers, M., Moor, I., Perelman, J., Richter, M., Rimpelä, A., et al. (2015). Smoking in school-aged adolescents: design of a social network survey in six european countries. *BMC research notes*, 8(1):91.

Lucchini, A. R. M. . E. (2002). La valutazione soggettiva ed oggettiva della disfonia: il protocollo sifel. In *presented at the Relazione ufficiale al XXXVI Congresso Nazionale della Societ Italiana di Foniatria e Logopedia*.

Martínez, D., Lleida, E., Ortega, A., Miguel, A., and Villalba, J. (2012). Voice pathology detection on the saarbruecken voice database with calibration and fusion of scores using multifocal toolkit. In *Advances in Speech and Language Technologies for Iberian Languages*, pages 99–109. Springer.

McCallum, S. (2012). Gamification and serious games for personalized health. *Stud Health Technol Inform*, 177:85–96.

McNamara, A. P. and Perry, C. K. (1994). Vocal abuse prevention practicesa national survey of school-based speech-language pathologists. *Language, Speech, and Hearing services in schools*, 25(2):105–111.

Miller, A. S., Cafazzo, J. A., and Seto, E. (2014). A game plan: Gamification design principles in mhealth applications for chronic disease management. *Health informatics journal*, page 1460458214537511.

Naylor, P., Kounoudes, A., Gudnason, J., Brookes, M., et al. (2007). Estimation of glottal closure instants in voiced speech using the dypsa algorithm. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(1):34–43.

Nerrière, E., Vercambre, M.-N., Gilbert, F., and Kovess-Masféty, V. (2009). Voice disorders and mental health in teachers: a cross-sectional nationwide study. *BMC Public Health*, 9(1):370.

Nicollas, R., Garrel, R., Ouaknine, M., Giovanni, A., Nazarian, B., and Triglia, J.-M. (2008). Normal voice in children between 6 and 12 years of age: database and nonlinear analysis. *Journal of voice*, 22(6):671–675.

Pinilla, J., Gonzalez, B., Barber, P., and Santana, Y. (2002). Smoking in young adolescents: an approach with multilevel discrete choice models. *Journal of epidemiology and community health*, 56(3):227–232.

Ross, M. J., Shaffer, H. L., Cohen, A., Freudberg, R., and Manley, H. J. (1974). Average magnitude difference function pitch extractor. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 22(5):353–362.

Simons-Morton, B., Haynie, D. L., Crump, A. D., Eitel, P., and Saylor, K. E. (2001). Peer and parent influences on smoking and drinking among early adolescents. *Health Education & Behavior*, 28(1):95–107.

Sun, X. (2002). Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 1, pages I–333. IEEE.

Talkin, D. (1995). A robust algorithm for pitch tracking (rapt). *Speech coding and synthesis*, 495:518.

Tan, L. and Karnjanadecha, M. (2003). Pitch detection algorithm: autocorrelation method and amdf.