Biomedical Signal Processing and Control 8 (2013) 799–810

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Technical note

Effective pre-processing of long term noisy audio recordings: An aid to clinical monitoring

S. Orlandi[a,b], P.H. Dejonckere[c,d], J. Schoentgen[e], J. Lebacq[f], N. Rruqja[a], C. Manfredi[a,*]

[a] Department of Information Engineering, Università degli Studi di Firenze, Firenze, Italy
[b] Department of Electric Energy and Information Engineering, Università degli Studi di Bologna, Bologna, Italy
[c] Catholic University of Leuven, Neurosciences, Exp. ORL, Belgium
[d] Federal Institute of Occupational Diseases, Brussels, Belgium
[e] Laboratories of Images, Signals and Telecommunication Devices, Faculty of Applied Sciences, Université Libre de Bruxelles, Brussels, Belgium
[f] Institute of Neuroscience CEMO, Université Catholique de Louvain, Avenue Hippocrate 55 bte B1.55.12, B-1200 Brussels, Belgium

ARTICLE INFO

Article history:
Received 20 April 2013
Received in revised form 30 July 2013
Accepted 31 July 2013
Available online 29 August 2013

Keywords:
Voice analysis
Biomedical signal processing
Long duration recordings
Otsu method
Voiced/unvoiced selection
Synthetic signals
Newborn infant cry

ABSTRACT

Nowadays, great attention is devoted to minimizing the discomfort caused by connection of patients to sensors for long-term monitoring of physiological parameters. Hence, the need for contact-less monitoring systems is increasingly recognized in clinical investigation. To this aim, audio signals recorded by ambient microphones are an appealing and increasing field of research: in the biomedical field, application of contact-less audio recording of long duration may concern obstructive apnoea syndrome, preterm newborns in Intensive Care Units, daily monitoring in occupational dysphonia, speech therapy, Parkinson and Alzheimer disease, monitoring of psychiatric and autistic subjects, etc. However, a significant amount of ambient noise is inevitably included in the records.

Especially in the case of recordings that take a long time, manual extraction of clinically useful information from a whole record is a time-consuming operator-dependent task, the length of a whole recording (even several hours) being prohibitive both for perceptual analysis made by listening to it and for visual inspection of signal patterns. Moreover, objective measures of signal characteristics may serve clinicians as a common ground for diagnosis. Hence, automatic methods are needed to speed up and objectify the analysis task.

The present work describes a new, automatic, fast and reliable method for extracting "voiced candidates" from audio recordings of long duration for both clinical and home applications.

To demonstrate its effectiveness, the method is compared to existing software tools commonly used in biomedical applications using synthetic signals.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Nowadays, great attention is devoted to minimizing discomfort for patients who repeatedly have to undergo long-term monitoring of physiological parameters that require spending long periods of time wearing several sensors connected by cable to a data acquisition system. Moreover, some categories of patients need special attention because of age or specific neurological/psychological conditions such as infants, very young children, adults and elderly patients suffering from depression, mental disorders, etc. Unobtrusive sound-recording and quantifying relevant acoustic parameters over time would enable the clinician to correctly understand and accurately pinpoint the moment and duration of inadequate/inappropriate behaviours. The availability of dedicated systems performing fast automatic audio signal analysis is thus a very promising perspective to overcome the above limits and enhance services for diagnosis and treatment in hospitals, during working activities and at home.

Hence, the need for contact-less monitoring systems is increasingly recognized in clinical settings. To this aim, audio signals recorded by contact-less ambient microphones are an appealing and increasing field of research, thanks to the growing technological developments that provide reliable and cost effective recording tools.

A considerable limitation of routine use of methods concerning recordings of long duration is the amount of time that conventional observation of records by the specialist is unacceptably time-consuming, "events" of interest (mainly voiced frames) being detected manually or with semi-automatic methods. Moreover, with some categories of patients, providing (quasi-) real time

* Corresponding author at: Department of Information Engineering, Università degli Studi di Firenze, Via S. Marta 3, 50139 Firenze, Italy. Tel.: +39 55 479 6410; fax: +39 55 4796497.
E-mail addresses: claudia.manfredi@unifi.it, manfredi@det.unifi.it (C. Manfredi).

1746-8094/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.bspc.2013.07.009

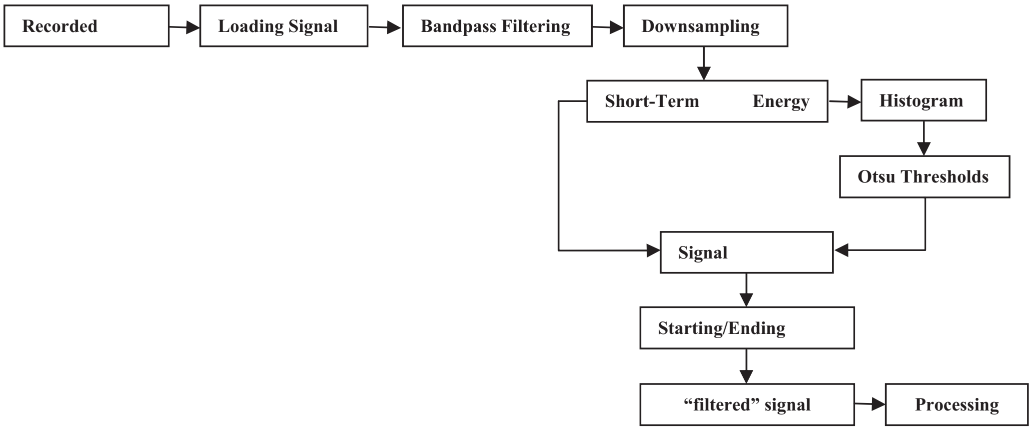

Fig. 1. Flow chart of the LTAA procedure.

feedback to the patient would be desirable as inadequate vocal behaviour patterns could be signalled and then, hopefully, corrected or reduced. Also, manual extraction of meaningful events, an operator-dependent task, is not fully reliable. To overcome these problems and avoid problems related to the use of microphones, some devices are commercially available [1–3], based on a contact sensor to be glued on the neck of the patient that, however, may cause discomfort. At the end of the recording session, signals must be downloaded into a computer to be analyzed off-line with specific software tools implementing techniques that (indirectly) extract relevant clinical parameters. The pre-processing step we propose here, though conceived for contact-less (microphone) audio recording, could perhaps speed-up the analysis with these tools for a first data screening and memory savings.

The analysis step should be preceded by fast and automatic extraction of meaningful "events", i.e. those intervals of the audio signal that may carry useful clinical information to be analyzed by means of several available signal processing techniques, according to the parameters of clinical interest in the specific application. In the case of long-duration recordings and/or quasi real-time requirements, this can be achieved if a reliable and fast procedure is implemented to "filter out" unwanted components. At present, this challenging problem that limits the application of most available software tools, especially in the case of prohibitively long recordings.

This work presents an automatic method for fast processing of long-duration audio recordings coming from the human vocal apparatus (i.e. sound is recorded at the lips) by means of a reliable procedure to be used both for clinical, occupational and home applications. It performs the selection of those intervals of the signal that may be candidates for further analysis (usually voiced intervals) without requiring a prior knowledge on the speaker's fundamental frequency nor any empirical threshold setting. This step may be followed by any specific analysis as required.

Contact-less long-duration of audio recording may concern: occupational dysphonia [4–10], speech therapy in Parkinson and Alzheimer disease [11,12], emotion and depression [13–15], autism [16,17], preterm newborn cry [18–21], obstructive apnoea syndrome [22–24].

To make appropriate settings on some parameters, the tool, named Long Term Audio Analyzer (LTAA), is equipped with a user-friendly interface that allows signal loading and parameters setting. In fact LTAA is a 'plastic' tool that can be programmed specifically depending on the type of signals considered as 'interesting' in a given application, thus extending the concept of 'voiced' to the concept of 'vibration' (occurring within the vocal apparatus).

Once the settings are made, the procedure is fully automatic and provides the "filtered" signal (i.e. the signal consisting of "sound" intervals only) in a few seconds. Tables and plots can be obtained summarizing the results. According to the application, LTAA may be followed by other steps for more detailed analysis and classification.

LTAA is compared to other existing voice analysis tools commonly used in the biomedical field [25–27] for effectiveness in extracting voiced intervals of the signal. The methods are tested on synthetic signals. A first set is provided by a voice synthesizer built by some of the authors of this paper [28,29] that was already used to create audio signals closely resembling adult voice [30–33]. A second set is based on a model obtained from newborn infant cry. Both of them will be described in Section 3.

The procedure is described in detail in Section 2. Section 3 gives experimental results. Concluding remarks are reported in Section 4.

2. Materials and methods

The aim of the proposed analysis system is the automatic detection of meaningful sound events from the whole audio signal. This is achieved by means of the following steps:

A. Pre-processing: band pass filtering and down sampling.
B. Automatic segmentation: detection of the sound intervals of the signal and signal "filtering".

To deal with quasi-stationarity, the analysis is performed on subsequent short signal frames. Possible electrical noise corrupting the signal is reduced in the implemented pre-processing step. Other remaining unwanted sounds are dealt with in the following step.

The proposed method was developed under Matlab r2012a package. A flow chart is shown in Fig. 1.

If required to work off-line, the proposed method, named Long Term Audio Analyzer (LTAA), is provided with a very simple user-friendly interface (Fig. 2) that allows for the following options:

- Load: the user loads the audio signal to be processed.
- Setup: the user sets the following parameters for subsequent processing.
  - o Sampling frequency (default 44.100 kHz).
  - o Down-sampling frequency (default 11.025 kHz).
  - o Starting and ending time instant, to select the part of the signal to be processed (default: whole length).

Download English Version:

https://daneshyari.com/en/article/10368443

Download Persian Version:

https://daneshyari.com/article/10368443

Daneshyari.com