App development with audio applications from m-file to app
Project 1

Daniel Jankovic

June 29, 2015

Voice Activity Detector (VAD) is a technique used in signal processing to detect the presence of human voice in a signal. It is basically an energy detector that indicates speech when the energy of the filtered signal exceeds a predifiend threshold. Considered an important technology in speech based communication, today there are various types of applications that use it. Therefore a wide variaty of VAD algorithms have been developed to provide the needed features.

There are different kind of stand-alone commercial baby monitors on the market today. From the most basic, that use one-way radio communication, to advance two-way communication monitors that use signal processing to transmitt audio when a predefined threashold has been reached. It is also possible to find baby video monitors that broadcast both audio and video when the sensors notice movement. Since the majority of monitor applications rely on radio signals to communicate between the units there is a risk that the signal strength will weaken or possibly not even reach the receiver becuase it needs to pass through multiple walls of varying thickness. The signal could also be effected by other applications. As the stand-alone monitor focuses on reliability (among other sale important strategies), little is known about the security features. It is possible to assume that the communication is uncrypted, at least in some products, and therefore introduces a potential risk for intrusion of peoples privacy.

To resolve the issuses broughten up above, an application such as the BAD can can be made more portable, versatile and secure with the help of todays smartphone technology and VAD. There are many VAD algorithms to choose from and they all have their strengths and weaknesses. Complex algorithms such as Linear Predictive coding (LPC), mel-frequency cepstrum (MFC) are very powerful but quite difficult to grasp and probably to implement, they can be considered out of scope for this course. The following algorithms are easy to implement and can be, when combined, quite robust for the task of a basic BAD. The simple short-time energy algorithm calculates the energy levels for each frame to detect voice, unvoiced or silenced regions. Voiced regions will have higher energy levels, however, the algorithm does not take unwanted noise into

account which means that we can have false indication of voice detection. In order to remove the noise from the signal, spectral subtraction can be preformed. In the case of BAD, the threshold needs to be adjusted so that unforeseeable sound is not intepret as the infants cry. Zero-crossing rate (ZCR), is the rate at which a signal changes from plus to minus and back. The higher the rate the higher the frequency which indicates possible voice activity. According to [1] the cry sound that an infant makes has a fundamental frequency of 250-600 Hz (pitch). To be able to use ZCR together with the information above, it is nessassary to extract the pitch from the signal in order to match the frequency interval. This can be made with cepstrum method? EXPLAIN!

The main task of BAD is to detect infant activity and an alternative algorithm is proposed in [1]. It describes an cry detection algorithm that is build up by three main stages. i) *VAD*, a statistical model-based detector is used for detecting sections with sufficient audio acitivity. Also it helps to reduce the power comsuption. ii) *Classification*, uses k-nearest neighbours (k-NN) algorithm to lable each fram as either 'cry' (1), close enough, or 'no cry' (0). iii) Post-processing, validation stage in order to reduce false-negative errors. The idea of having devoted algorithm to detect infant cry is a winning concept for a BAD application, according to the authors it even had promising results in low SNR. Despite simplicity of the algorithm many of the features required to implement were mentioned earlier to be out of scope for this project.

An algorithm that might be of interest [3] suggests an alternative way of speech enhancement without VAD technology. The singal is devided into multiple subbands and an noise floor level estimate is calculated simultaneously as the short-time average. The goal is to boost the subbands with high Signal-to-Noise Ration (SNR) rather than to supress the lower. This algorithm has great potential to reduce the noise levels when analyzing incoming signals to the BAD application.

++Part where I analyze open source algorithm and existing app++

A porposed algorithm for a BAD application

```
Get frame from the register
Devide the signal into different subbands
Calculate the total short-time energy average
Calculate noise level for each subbands
Calculate the gain for each subband
if energy average above threshold
    Boost the subband with high SNR
    Count the ZCR under 2 seconds
        if the ZCR is within 250-600 Hz (for 2 sec worh of frames)
            Possible infant activity detected!
            (Occurs only first time, and needs to be reseted)
            Record the sound for 10 seconds or until rates drops
            Send the recording to the receiver
        end
else
```

```
Reset ZCR
Dismiss and get next frame from register
```

References

[1] R. Cohen, Y. Lavner, *Infant Cry Analysis and Detection*, 2012

[2] E. Verteletskaya, K. Sakhnov *Voice Activity Detection for Speech Enhancement Applications*, 2010

[3] N. Westerlund, M. Dahl *Speech Enhancement using an Adaptive Gain Equalizer* 2003

[4] R. Narayanam *An Efficient Peak Valley Detection based VAD Algorithm for Robust Detection of Speech Auditory Brainstem Responses* 2013