

**A survey of voice detection and recognition algorithms**Suraj Mane¹, Aaditya Shah², Nirmitt Shrivastava³, Ankit Srivastava⁴, Prof. Bhushan Thakare⁵^{1,2,3,4,5}Computer Department, Sinhgad Academy Of Engineering

Abstract This paper provides a brief insight of some famous and particularly important algorithm used for voice detection. Voice detection is a technology used by computer system to detect speeches. Voice activity detection VAD is algorithm used to detect and recognise voice/speech.

Introduction

Pitch of the human voice is one in all and most simply and easy to control acoustic attributes. It plays an important role in each the production and perception of speech. Automatic speech detection uses the method Associate in Nursing the required technology for modifying speech signals into a sequence of words or different linguistic units and element by means that of an algorithmic rule enforced as a worm. Speech recognition is a technique that enables a device to recognize and understand spoken words, by digitizing the sound and matching its pattern against the stored patterns. Speech is the most natural form of human communication and speech processing has been one of the most exciting areas of the signal processing. Human machine interface system supported speech attracts a lot of interest supporting the fast improvement of the main frame performance, an effective VAD methodology out there for these non-stationary interferences, victimisation their high performance speech action system" AMNOR (Adaptive electro-acoustic transducer array for noise reduction)". He uses electro-acoustic transducer array to discriminate signals utilizing direction distinction between speech and interference. This paper explores about different rising speech detections hardness in adverse environment voice policy investigation methodology that has blessing in noise surrounding by victimisation speech maximization and silence feature standardisation.

Algorithm**Voice Activity Detection (VAD):**

VAD was proposed in 1980 by Boll Steven, VAD is an algorithm which detects human speech and recognise human voice. Main purpose is to identify discrimination of speech from silence or other background noise. The VAD algorithm is based on any other combination of general speech properties, such as temporal energy variation, periodicity, and spectrum. VAD indicates the presence or absence of speech as observed by Ramirez 2004.

Basic principle of VAD device is that it extract some measured features or quantities from the input signal and then compares these value with thresholds. Voice activity is declared if the measured value exceeds the thresholds. Otherwise is declared for no speech activity.

FEATURES:

- (i) Good decision rule.
- (ii) Adaptability to background noise
- (iii) Low computational complexity

Long-term spectral flatness measure:

Spectral flatness is measure of the width, uniformity and noisiness of the power spectrum. A high spectral flatness indicates that the spectrum has a similar amount of power in all spectral banks, and the graphs of the spectrum would appear relatively flat and smooth. This would sound similar to white noise. A low spectral flatness indicates that the spectral power is less uniform in frequency structure.

This long term measure is computed victimisation the spectra of the last R frames of the signalling $x(n)$. The this feature, $L_x(m)$, at the MTh frame and across all the chosen frequency is then calculated by dividing the mean value of the ability spectrum by the expected value of the ability spectrum to expand the dynamic vary, its measured on a graduated table, starting from zero to minus eternity.

Spectrum normalization (SN)

Comparing spectrum in several measure standards could be troublesome once examination the variations between different speeches signals. Hence, the employment of normalization could maintain the measures commonplace. In different words, the normalization reduces the error once examination the spectrum. Before analysing the spectrum variations for different words, the primary step a normalize the spectrum by the linear normalization. When normalization, the values of the spectrum are obtained or normalized spectrum is about to the interval [0,1]. Normalization changes the worth vary of the values of FTT, consequent step in programming the speech recognition is to

look at spectrum of the recorded speech signals. Then the algorithm are compared supported the differences between the check or target signals and therefore the training signals or references signals.

Voice Activity Dectector:

Voice Activity Detector by SPIRIT Corp. was deveoped for integrating into various smart system, this voice activity detector detects the avaiability of voice/speech in signals and even provide vigorous performance in noisy atmosphere. This voice activity detector has ability to suits itself for any particular application. XDAIS framework is provided by SPIRIT corp. to develop application using this algorithm.few framework functions

ALG_create

- For memory allocation/algorithm instance creation

-

ALG_delete

- For memory de-allocation/algorithm instance deletion

-

ALG_activate

- For instance activation

ALG_deactivate

- For instance de-activation

-

ALG_init

- For instance initialization

-

ALG_exit

- For instance exit operations

-

ALG_control

- For instance control operations.

the fundamental thought of a speech recognition system is intended to perform 2 operations. These operations area unit the modelling of the signal underneath investigation and matching of patterns mistreatment the extracted options obtained from the primary step. Signals modelling arms to search out a group of parameters from the conversion of the speech signal. The task is to search out parameter set from previously processed memory with matching properties of the parameter set obtained from the target speech signal to certain extent that represents the pattern matching step.

Wiener filter (WF): This filter used for recognizing speech

FIR Wiener filter is used to estimate the specified signal fromthe observation methodto induce the calculable signal. The desired or check and train signals are assumed to be correlated and bring together wide-sense stationary. The aim of wiener filter is to decide on the acceptable filter order and to find the filter coefficients and with that the system will get the best estimation. In different word the system will minimize the mean-square error with correct coefficient. The recorded signals are wide-sense stationary processes. Then, the references or training signals will be used as input signals, and recorded check signals will be used as desired signals.

Hidden markov model (HMM)

A hidden markov model could be triple (π, A, B) where each chance within the state transition matrix and emission matrix is time freelance, which implies that the matrices don't amendment in time because the system evolves. In practice, this model is one in all the foremost delusive assumptions of Andre Markov models concerning real processes HMMs, represented by a vector and 2 matrices(A,B), that square measure if nice worth in describing real systems. Although, the model is typically associate degree approximation, the results square measure amenable to analysis. The commonly resolved issues square measure.

LTSF: The proposed long term spectral flatness method based VAD algorithm:

Long-term: the suggested VAD algorithmic rule assumes that the signal spectrum is a lot of organised throughout speech segment than throughout noise segments. It adopts the typical spectrum over a semi-permanent window rather than instant values of the spectrum. Typically, a period gram is usually used for spectrum estimation, however its documented that the period gram is associate in nursing inconsistent spectral computer as per, the Welch-Bartlett technique was found to grant an honest trade-off between variance reduction and spectral resolution reduction. Therefore, in their projected algorithmic rule, the signal spectrum is calculated mistreatment the Welch-Bartlett technique.

A LONG TERM SPECTAL FLATNESS MEASURE- based VAD algorithmic rule is that the input speech signal is rotten into frames of twenty Ms long with associates in nursing overlap of ten mili-sec by the Hann window. The spectrum of the segmental signal is calculable mistreatment the Welch-Bartlett technique. At the MTh frame, the LONG TEM SPECTAL FLATNESS MEASURE feature $I_x(m)$ is computed mistreatment the previous R frames. The initial call regarding whether or not there contains speech within the last R frames is created through the comparison with associate

in nursing adaptive threshold. The initial call is denoted by VINL, (m)=1; otherwise, VINL(m)=0 and there are solely non-speech frames over the previous R frames. They have a tendency to adopt the option theme projected by Ghosh at to create the ultimate VAD choices in a 10-ms interval. Then, the target 10-ms interval. First, the initial choice, VINL (m), VINL (m+1), VINL (m+r-1), are collected for those long interval. Then, the target 10-ms interval is marked to be speech, if there is eightieth or higher than of these initial choices that contain speech, otherwise, it's marked as non-speech. The eightieth was gotten by trial and error that provided the most VAD accuracy for many noises tested over 5 SNR levels spectral flatness live.

Conclusion

Here we have tried to survey few voice detection algorithms in a simplest way. And understand the working of each algorithm which enables us to get an idea about voice detection techniques.

References

- Rabiner LR, Sambur MR: **An algorithm for determining the endpoints of isolated utterances.** *Bell Syst. Techn. J* 1975,54
- Ghosh P, Tsiartas A, Narayanan S: **Robust voice activity detection using long-term signal variability.** *Audio, Speech, and Lang. Proc. IEEE Trans* 2011,19
- Prasad RV, Sangwan A, Jamadagni HS, Chiranth MC, Sah R, Gaurav V: **Comparison of voice activity detection algorithms for VoIP.** In *Proceedings of the Seventh International Symposium on Computers and Communications (ISCC'02), Washington, DC.* Piscataway: IEEE; 2002:530–535.
- Davis A, Nordholm S, Togneri R: **Statistical voice activity detection using low-variance spectrum estimation and an adaptive threshold.** *Audio, Speech, Lang. Proc. IEEE Trans* 2006,14(2):412–424.
- Madhu N: **Note on measures for spectral flatness.** *Electron. Lett* 2009,45(23):1195–1196
- Khoa PC: **Noise robust voice activity detection.** Master's thesis, NangYang Technological University, 2012
- Deepak, M. Vikas, "Speech Recognition using FIR Wiener Filter", *International Journal of Application or Innovation in Engineering & management (IJAIEEM)*,pp.204-20,2013.
- Paul, Teenu Therese, and Shiju George. "Voice recognition based secure android model for inputting smear test result." *International Journal of Engineering Sciences & Emerging Technologies*, ISSN: 2231-6604,2013.
- Mohan, Bhadrageerajagan, and N. R. Babu. "Speech recognition using MFCC and DTW." *Advances in Electrical Engineering (ICAEE), 2014 International Conference on.* IEEE, 2014.