Voice to Command Converter

Akshay J. Sonawane[1] Aniket D. Dalvi[2] Tushar P. Gade[3] Gokul T. Dhurjad[4]

[1,2,3,4]Department of Computer Engineering

[1,2,3,4]Shatabdi Institute of Engg. & Research, Agaskhind Nashik, Maharashtra, India

Abstract— In this paper, we tend to represent a brand new technique for sturdy Voice to command conversion technique. Voice to command conversion may be a method by that system is ready to grasp linguistic communication. By the assistance of projected technique we'll be ready to offer computing system the voice command to perform task. the primary step is to require one audio as associate input. The second step is to eliminate the presence of noise. The third step is to match the patterns to acknowledge the input audio. The error tolerance mechanism is enforced so it may be sturdy against human errors, setting noise. there's no demand of storing the audio for matching, that results in decrease the area and time complexness each compare to alternative systems. The information utilized in the experiment can solely be consisting of varied voice patterns. Patterns ar onerous coded in such some way that system isn't required to be trained by the tip users.

Key words: Voice Recognition; Natural Language Processing; Fast Fourier Transform; Hidden Markov Model

I. Introduction

With the rapid climb of the digital world, speech recognition is turning into in style. it's conjointly become a preferred research topic. The voice because the interface is turning into progressively acceptable for the devices. Speech recognition comes below the branch of science referred to as tongue process (NLP) that is that the field of applied science, computer science and linguistics that area unit involved with the interaction between tongue (spoken by humans) and pc systems. Informatics allows the pc system to drive, which means out of the input language. It makes a lot of less complicated for the user to control the device. Recent analysis proposal a general technique, that acknowledge the voice by coding it to patterns, These patterns area unit used for the regulation of weighted automata that generate an additional worth that decides the chance of a specific word is spoken or not. A future quantity of labor is completed on the speech recognition system to cut back issues like noise, mismatches within the audio signal, that area unit caused thanks to the variability in inter/intra speaker, totally different accent of chatting with win the performance hardiness. Within the planned technique we've got used LRU technique which period needed for looking pattern within the info. during this technique we tend to keep the pattern rather than the voice that once more decreases the area quality of the system. The structure of paper is as follows: In section II we tend to provides a temporary introduction of speech recognition system and also the info used for the system [2]. In section III we'll be discussing however voice signals, the noise is declared and existing strategies. In section IV we'll propose our technique and compare results. The results area unit summarized in section V.

II. Voice to Command System and Database

The speech recognition system is continuous mixture density hidden Andre Markoff model (HMM) system whose parameters area unit calculable by Viterbi coaching [2]. In our Speech recognizing technique initial the standardization is completed, that is to search out Cepstral constant by taking the Fourier rework of a brief time window speech followed by decorating the spectrum victimization inverse fourier.

Fourier reworks and so acknowledges the pattern for various speakers and recording conditions. during a second step phone area unit recognized on the idea of the pattern.

The info are going to be consisting of a special set of patterns which is able to represent the states needed for the undefeated completion for a given phone. this can have dominance over the opposite ways that use warehouse of coaching audio, for quite just one occasion. The structure checking of audio input from the required audio is completed. The info additionally contents weight, assign to every state for the undefeated completion, input voice needs to complete and cross through the brink limit. If any input audio doesn't with success complete stages of any phone within the info, then its output is directed to the info phone whose threshold limit is near threshold limit of the input phone (audio). looking out of the phone within the info system on the technique within which last used voice patterns area unit unbroken at the highest within the info. this system is popularly called LRU (Least Recently Used). the most effective a part of victimization this system is to decrease the time quality at the time of looking out the required phone. this system makes it additional economical than others. the downside of standard voice classification supported phones is that sub-word units selected for the aim of classification don't seem to be reliable as a result of comparatively low phoneme recognition rate [3].

III. Notations and Assumptions

Several ways mix the enframing the composition with a style of the pattern supported the FFT of the signal [6]. whereas the energy of the full signal is preserved by such a metamorphosis, the computed energy on every interval is also drastically modified [7]. we have a tendency to use the notation F (x) denoting Fourier remodel Associate in Nursing Denoting inverse Fourier remodel of an audio signal. we have a tendency to initial took the Fourier remodel F (x) of the input audio signal. Its information is extracted victimization Fourier remodel, that information is discretion victimization inverse Fourier remodel. N (omega) is taken into account as a standardization perform holding noise eliminated values..

A. Noise Elimination

We assume that the discovered signal could be a realization of wide sense stationary method [4]. If the signal to noise quantitative relation isn't too low, an easy methodology to sight speech relies on signal energy [2]. Because the level of energy within the speech signal is over the amount of noise energy. On this basis a threshold limit are often obligatory on the energy signal, all the values higher than it'll be thought-about as vocalization, data and every one the worth’s below the edge value are discarded as likelihood those values are nice.

In F (x) >threshold limit

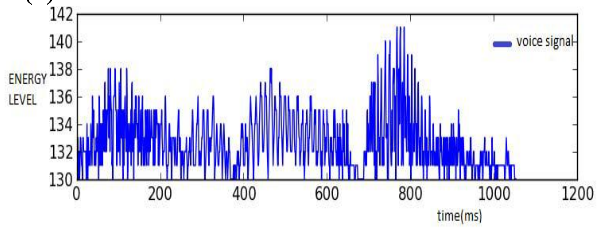

Fig. 1: A graph showing energy level after Noise Elimination

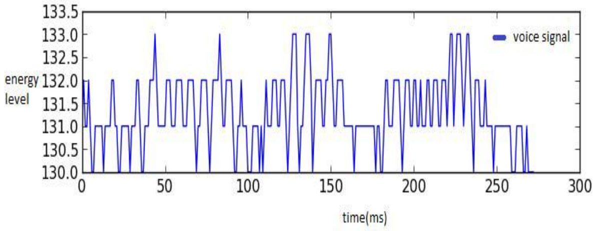

Fig. 2: Graph

IV. Proposed Method

The motivation behind this technique is to provide voice commands to the system by creating a speech auditory communication. As currently within the digital world mistreatment voice command to work a tool are going to be a lot of appreciated and a lot of convenient than mistreatment mouse. As mentioned in Fig a pair of, initial audio is taken from the mike that could be a linguistic communication with the environmental noise. Within the second step, audio file generated by the sound recorded from the mike, is employed to get Fourier remodel of it which supplies the sound energy state knowledge on the time. Within the third step, the noise elimination is finished on the information generated by the Fourier remodel. this is often to form make sure that the audio is currently free from noise, that intern increase the probabilities of coding the most voice command from the audio file recorded for the popularity. within the fourth step, noise eliminated knowledge is born-again into a string by finding the code transformation for every knowledge generated that is employed to cipher the voice to string pattern for extraction of the specified patterns. Within the fifth step, the patterns square measure matched from the patterns already keep within the info. the share of pattern matched offers a worth that is calculated by rundown the load related to the states of the automata, that should cross threshold worth for any word within the info for its roaring recognition. When the brink worth is crossed with success the task related to that command is finished. The task is additionally kept within the info like the word spoken.

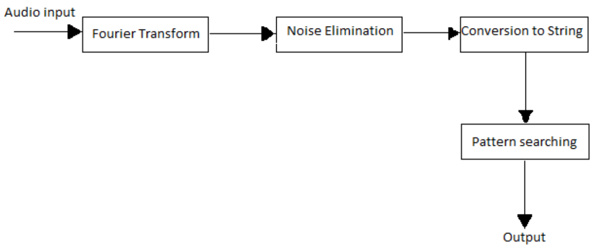

Fig. 2: Framework of Proposed method

Figure 3 below represents pattern to be followed by the utterance of word “home”. Each state has a weight to assign to it which is calculated by with the analysis of the pattern generated by the utterance of the word “home”.

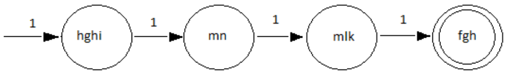

Fig. 3: State diagram of utterance of the word “home”.

As the states square measure achieved corresponding weights square measure further to a variable, which can be used at the top for comparison of the edge worth. This threshold worth can decide whether or not its auditory communication of a specific word or not. To cut back the time quality we have a tendency to store the word with the patterns within the information within the order of most often used words. this can increase the chance of obtaining the required output at the start of the information itself, instead of looking the full information. during this approach, we've got a plus over the opposite techniques that wont to train the system for the various user. we've got tested this methodology on males moreover as feminine voice and that we got accuracy of approx ninetieth. The ten quality thanks to distinction within the auditory communication of same word at totally different instance of your time. The area quality is additionally reduced because the information is storing the pattern of the words that square measure principally occurring to all or any the users. this can scale back the area, that was earlier used for storing the coaching voices for various users.

V. Results

We first took the convolution between the speeches of the same kind and the graphs generated by convolution gave us the result, that there are some common areas where the data of two or more utterances of the same word is same. For example, in Fig-4 we get common region 0-400 (approx taken along time axis having a same corresponding energy level), it ensures that if the pattern of the waves is generated in this region can be used as a state in the automata of the word “open”. Fig-4, Fig-5, Fig-6, Fig-7, Fig-8, Fig -9 convey the same.

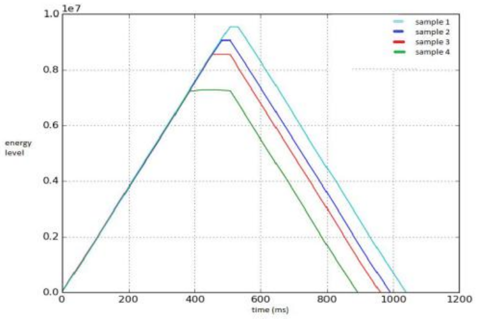

Fig. 4: Graph representing utterance of the word “Open”

In fig -5 there are plots of 4 utterances of the word "Google". It Indicates the utterance level of the word is at the same level on a time scale of 0-400 ms, corresponding with unique energy values, which indicates that patterns for "Google" will be found unique between these intervals compared to other words.

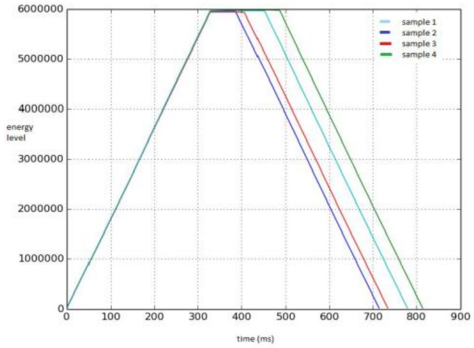

Fig. 5: Graph representing utterance of the word "Open"

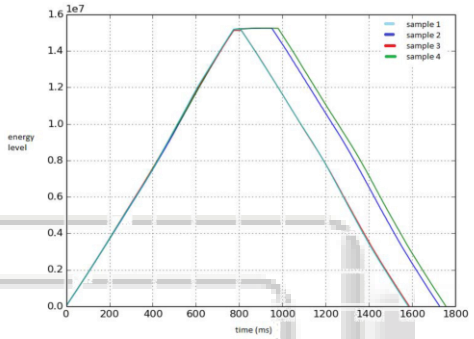

Fig. 6: Graph representing utterance of the word "Logout"

Figure 6 shows that the patterns for the word "Logout" can be found in the range 0-1000 (along the time axis).

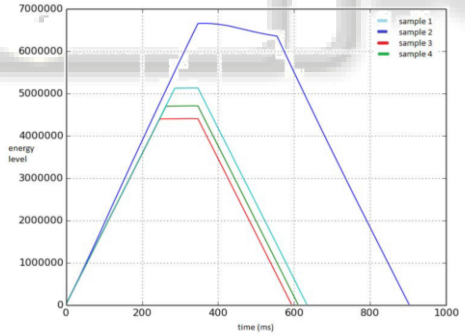

Fig. 7: Graph representing utterance of the word "Mute"

Figure 7 shows that the pattern for the word "Mute" can be found in the range 0-400 (along the time axis).

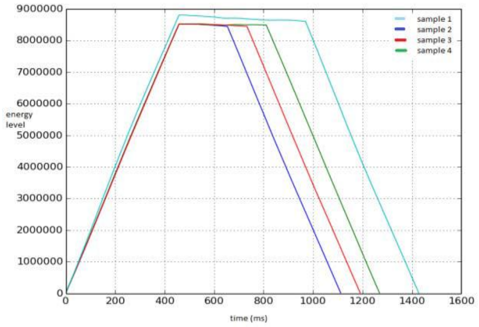

Fig. 8: Graph representing utterance of the word "Power off"

Figure 8 shows that the pattern for the word "Power off" can be found in the range 0-800 (along the time axis).

These graphs show that these common same regions wherever waves show similar plots, such as these regional patterns is generated which may be used because the state of the automata, that is additional used for the winning recognition of a voice. to acknowledge patterns we tend to 1st took the Fourier remodel followed by noise elimination. When noise elimination, knowledge is born-again into the string and therefore the patterns ar searched that ar same for constant quite words. The generated pattern for the word home is often seen in Fig three.

It shows the incidence of patterns on the time axis in voice signal of the house. If any voice follows these automata and crosses the brink worth, then that voice signal is also of word-"home". Accuracy of the planned technique is ninetieth, which is applicable for the feminine voices moreover as male. this system is nice for style a voice commanding system. Out of fifty utterances it's able to acknowledge forty five voice commands.

VI. Conclusion

In this paper, we tend to given the technique for the sturdy recognition of the word (voice command) auditory communication. it's a procedural approach during which voice has got to submit to the automata, gain total worth that should be bigger than adequate to the brink worth for no-hit recognition and applies error tolerance to attain lustiness against human errors. this method is suited coming up with voice commanding system. exploitation pattern for the entire recognition of a word, followed by threshold worth checking makes it a lot of sturdy to spot the auditory communication of the word accurately. 90%of the auditory communication is recognized properly. additional work are that specialize in creating a system a lot of economical and user friendly. we tend to square measure implementing this methodology on a voice commanding system.

References

[1] Changxue Ma, Uniterm Voice indexing and search for mobile devices, applications & software research center Motorola Inc 1295 e. Algonquin IL 60196.

[2] Volker Stahl, Alexander Fischer and Rolf Bippus, quantile based noise estimation for spectral subtraction and wiener filtering, Philips research laboratories.

[3] Olivier Siohan, Michiel Bacchiani, Fast Vocabulary-Independent Audio Search Using Path-Based Graph Index. INTERSPEECH, 2005.

[4] M. H. Hayes, "Statistical Digital Signal Processing and Modeling" John Wiley & Sons, Inc., 1996.

[5] C. Allauzen, M. Mohri & M. Saraclar. General Indexation Of Weighted Autometa – Application to Spoken Utterance Retrieval ACL, HLT, 2004.

[6] Jerome Lebose, Luc Brun, Jean Cluade Pailles, "A Robust Audio Fingerprint Extraction Algorithm".

[7] S. Mallat. A Wavelet tour of signal processing. Academic press, 1999. Chapter VIII p.363.

[8] S. Furui, "Recent advances in robust speech recognition," in Proc.ESCA-NATO Tutorial and Research Workshop on Robust Speech Recognition for

Unknown Communication Channels, Pont-a-Mousson, France, Apr. 1997, pp. 11–20.

[9] Y.-F.Gong, "Speech recognition in noisy environments: A survey,"Speech Commun., vol. 16, pp. 261–291, 1995.

[10] B.-H. Juang, "Speech recognition in adverse environments," Comput. Speech Lang., vol. 5, pp. 275–294, 1991.

[11] C.-H. Lee, "On feature and model compensation approach to robust speech recognition," in Proc. ESCA-NATO Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels, Pont-a-Mousson, France, Apr. 1997, pp. 45–54.