International Journal of Engineering & Technology, 7 (2.32) (2018) 400-404

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Conversion of Body Conducted Unvoiced Speech(Murmur) to Normal Speech Using Hidden Markov Model (HMM)

T.RajeshKumar[1], M.Srinagamani[2], M.Sai ram chandu[3], S.Mounika[4]

[1,2,3,4]*Department of CSE, K L E F, Guntur, India*
**Corresponding author E-mail: t.rajesh61074@gmail.com*

Abstract

The main purpose of this paper is Conversion of non- audible murmured voice into the normal speech using Hidden Markov Model(HMM).This non audible murmur voice NAM is a one type of murmured voice which can be delivered by a NAM microphone which is attached behind the speaker's ear. The Hidden Markov Models(HMMs) are stochastic models of statistical learning .These are very useful in speech recognition .The point of the paper is to collect as much as data from the device and convert it into audible and clear data signal that can be used for further sensory based applications. Hence, having an insight of how to convert the NAM to speech and then to whisper has a lot of benefits while keeping in mind the disadvantages of such conversion. Since, NAM is minute details of a communication between one's own self it is highly recommended to the data in as much as discrete format as necessary since a speech signal can have various frequencies over a portion of the signal, big data approach is recommended.

Keywords:*Non audible murmur, NAM microphone, Hidden Markov Model, Speech Signals.*

1. Introduction

Speech Communication plays a exceptionally crucial part in day by day life. It is the most well known strategy of Human Communication. In later years, the fashion of speech communication has impressively altered with the progress of innovation. For occasion, the hazardous distribution of mobiles has empowered individuals to exchange with each other and where ever they need and brought a more helpful fashion of talking to us.

In spite of the fact that mobiles made speech conversation likely to happen in different circumstances, there are really a few occasions where we encounter troubles in speech communication. For illustration, we can consider convenience of talking secretly in a large group of things like crowd, talking it self would now and then bother others in calm situations such as in library. We may loose our voice on the off chance that exposed to operation to expel discourse tissues like the pharynx because of pharyngeal cancer. Various impediments are as yet accessible in discourse correspondence. The change of advances to overpower these trademark issues of discourse correspondence is fundamental to make our discourse correspondence significantly more comprehensive.

As of late, quiet speech interfacing have pulled in consideration as a innovation to back unused discourse correspondence styles[1]. They enable discourse correspondence to take put without the need of transmitting an capable of being heard auditory signals. Few attempts are there to investigate detecting gadgets as choices to the air-conductive microphone like throat microphone[3], ultrasound imaging[5],EMG(electromyography)[4], and etc. Detecting gadgets are valuable to identify delicate speech in a personal discussion. More over holding attention as a talking help for the people who are vocally disabled. In expansion, those people are more viable for commotion solid discourse flag. For event, Subramanian et al[6]. are detailed that bone conducted discourse signals can be utilized more effectively to improve speech sounds under overwhelming conditions of noise.

NAM microphone is the one of the microphone which is identify the body directed discourse, the NAM mouthpiece is made by Nakajima et al.[7]. Moved by a stethoscope, the NAM receiver was initially created to distinguish greatly delicate murmured voice called NAM, which is so calm that individuals near by the speaker scarcely listen its radiated noise. In spite of the fact that NAM is a really diverse way between extremes from characteristic voices, it can be utilized effectively by anybody whose vocal body parts work sensibly good. Set on the neck under the ear, the NAM beneficiary is fit for seeing discuss vibrations in the vocal tract from the skin through as they say the delicate tissues of the head Qualified body-conductive narrative of different sorts of talk, for example, uncommonly fragile murmur as NAM, sensitive voices, low voice, and common discourse, is conceivable from any position since the conduction through pieces, for example, bones whose acoustic impedance is various from that of fragile tissues, is maintained a strategic distance from as one of the microphones to distinguish was initially created to identify greatly delicate murmur called NAM, which is so calm that individuals around the speaker scarcely listen its transmitted sound. Despite the way that NAM is extremely unmistakable medium from essential voices, it can be utilized satisfactorily by any one whose talk organs work commendably. It is what's more strong against outside voice inferable from its hullabaloo confirmation structure like in other body conductive microphones[9] . Subsequently, the NAM amplifier licenses us to discussion using diverse sorts of body-coordinated talk dependent upon the condition, e.g., non fit for being heard mutter or a body-drove whispered voice for calm talk correspondence and body led conventional discourse for commotion hearty

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

discourse correspondence Besides, its comfort is route superior to those devices, for example, ultrasound frameworks or EMG.

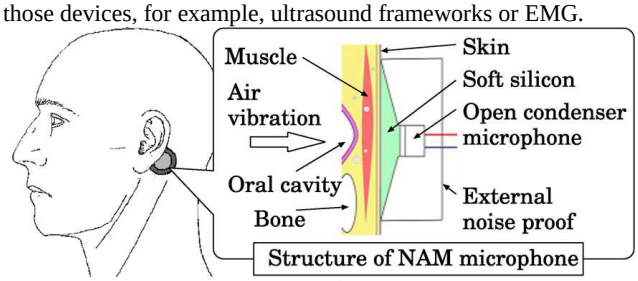


Figure 1. Structure and positioning of NAM(non -audible murmur) microphone.

NAM(Non audible murmured speech) AND BCW(Body conducted whispered voice): Non audible murmur speech BCW,NAM were concentrated on tupes of frame-directed voiceless correspondence . NAM is described as the verbalized age of archive calmers beyond utilizing the vocal-portage oscillation, passed on through the advancements and sagacious of talk body parts like tongue, feeling of taste, and besides lips, that wil be driven over in a way the delicate tissues of the head with no piece like bones[7] . NAM is recorded for utilizing the NAM enhancer related with the skin surface behind the ear, as showed up in Figure.1.NAM is a especially delicate whispered tone, Those control from claiming which will be Additionally little to anybody who would around a speaker on its transmitted sound. Really side, BCW may be portrayed Likewise A quiet tone drove over the delicate tissues of the head. We can converse with a predetermined amount of people abutting utilizing An quiet tone. since it controls satisfactorily gigantic to be equipped for being heard for them.

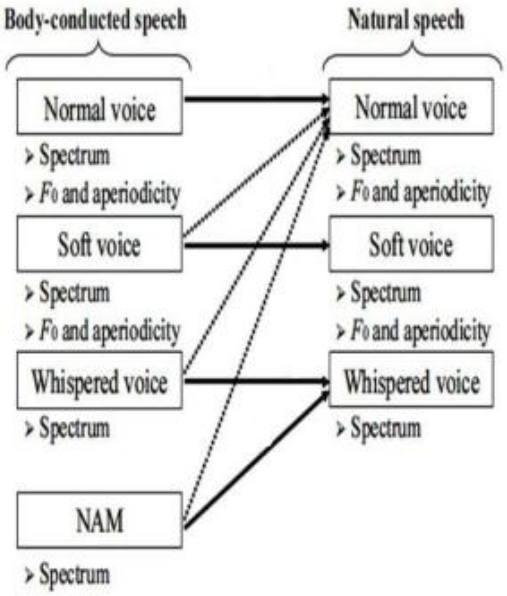


Figure 2.Body conducted speech conversion types

2. Literature Survey

In 2009 Tomkoi Toda and Steve Youthful has given a paper named direction preparing considering worldwide change for HMM-based discourse union. This paper has depicted another heading planning procedure beneath a necessity on a around the globe change (GV) for HMM based talk mix. The proposed procedure gives a headed together framework for getting ready and integrating talk using a typical premise, it yields incredibly basic headways in intuitive nature, also, it permits an all the more effective parameter period handle considering the GV in light of a close design methodology. Our following stage is to explore whether the proposed framework causes fundamental quality changes in composed talk differentiated and the ordinary GV based parameter age.

In 2010 Yamato Ohtani and Tomkoi Toda has given a paper named non parallel preparing for numerous to numerous Eigen voice change. [10]This paper depicted the EV GMM preparing strategy utilizing nonparallel information sets for some to-numerous EVC. In the proposed planning technique, a starting EV GMM is readied using parallel data sets involving a solitary reference speaker and various pre-put away speakers. By then, the beginning EV GMM is advance refined using non-parallel informational indexes involving a more prominent number of pre put away speakers while considering the reference speaker's voices as secured up factors. The exploratory comes to fruition have outlined that the proposed getting ready technique yields essential quality upgrades in changed over talk by effectively using non-parallel data sets checking a greater number of pre put away speakers.

In 2016 “Efficient Acoustic Commotion Cancellation In Non Capable of being heard Mumble Utilizing Wavelet Transform”. In this paper, they discuss about the measurable approach to upgrade body conducted voiceless discourse for quiet discourse communication utilizing wavelet change. A body-conducted voiceless discourse is called non-audible mumble, NAM receiver is viably utilized to distinguish exceptionally delicate voiceless discourse which is radiated exterior nearly unintelligible. Examination of NAM discourse has been made utilizing covered up Markov demonstrate (Gee) and Gaussian blend demonstrate (GMM). In this paper consider of analyzing NAM discourse utilizing haar and db2 as it is utilized to extricate the highlights of different sorts of discourse flag. Wavelet change is able of uncovering viewpoints of information that other discourse flag examination method such the extricate highlights are at that point passed to classifier for the acknowledgment of discourse. The test comes about appear that NAM is successfully changed over to Ordinary Discourse which progresses intelligibility.
we may encounter trouble in talking beneath honest to goodness uproarious conditions.[2] Also we may questionable to furtively conversation in a bunch and talking itself would at a few point offers inconvenience to others in calm circumstances, for illustration, in a haven, library. The advance of propels to prevail these in born in comforts of talk correspondence is essential to extend our talk correspondence. One of the sensible techniques to give them is to perceive talk sound through body-conduction using body conductive receivers.

The unprecedented employments of body conductive talk alter are in the field of security zone, Central division of Examination, while investigating reality from aggressors or criminals at detainment facilities, in the wake of applying difficult strategy for examination, at long final they take an instrument of implanting medicine on their body or hypnotize them and bring out the data. The wrongdoers mutter at blacked out state. At this organize, a NAM recipient and Wi-Fi transducers are joined behind the ear. At war field, the rules might be passed on by commando to their troopers silently with the utilization of this NAM sensors and farther correspondence frameworks. mouthpiece, one of the standard body-conductive authorities, is basically all the more overpowering against outer clam or differentiated and a standard discuss conductive intensifier. In talk correspondence we more frequently than not utilize distinctive sorts.

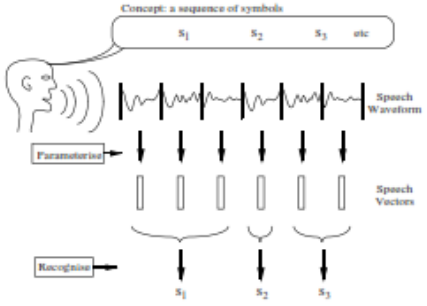
In 2006 “Unvoiced speech recognition using Tissue conductive acoustic sensor” they displayed non capable of being heard mumble acknowledgment in clean and loud situations utilizing NAM mouthpieces.[8] A NAM receiver is a uncommon acoustic gadget connected behind the talker’s ear, which can capture exceptionally discreetly expressed discourse. Non capable of being heard mumble acknowledgment can be utilized when privacy in human-

machine communication is wanted. Since non capable of being heard mumble is captured specifically from the body, itis less touchy to natural commotions. To appear this, we carried out tests utilizing recreated and genuine boisterous information. Utilizing mimicked boisterous information at 50 dBA and 60 dBA commotion levels, the non capable of being heard mumble acknowledgment execution was nearly rise to that of the clean case. Utilizing, in any case, information recorded in loud situations, the execution decreased.

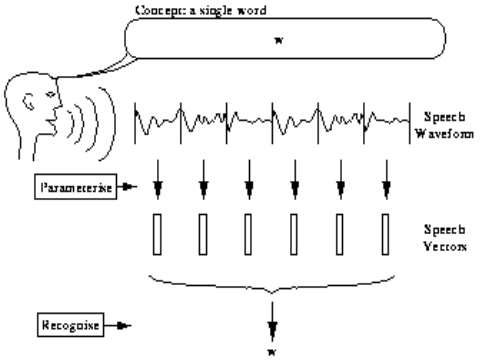
3. Methodology

Hidden Markov Models (HMMs)

It is utilized to show any type of time arrangement and the centre of HTK is additionally common reason. HTK is a toolkit for architecture Covered up Hidden Markov Models (HMMs). In any case, HTK is basically planned for building HMM based discourse handling apparatuses, in specific recognizer. Hence, much of the foundation back in HTK is committed to this task. There are two main handling steps included. Firstly, HTK preparing apparatuses are utilized to guage the parameters of the HMMs set utilizing preparing expressions and their related translations. Besides, unknown expressions are translated utilizing the HTK acknowledgment instruments. In any case, sometime recently propelling into detail it is essential to get it a few of the essential HMMs standards. It is moreover Supportive to have an outline of the toolkit and to have a few acknowledgment of how preparing and acknowledgment HTK is composed Common standards of HMM



Figure(3.1): Message Encoding/Decoding



Figure(3.2): Isolated Word Problem

Speech acknowledgment frameworks for the most part expect that the discourse flag is an acknowledgment of a few message encoded as a arrangement of one/more images (Figure 3.1). Yo effect the invert method of perceiving the basic image grouping given a talked expression, the persistent discourse waveform is first changed over to an arrangement of similarly separated distinct framework direction. This procedure of framework direction is perceived to plot a right delineation of the discourse waveform on the begin that for the term secured by a solitary vector (commonly 10ms or close). The talk waveform is seen as being static. In spite of the fact that this is not entirely genuine, it is a sensible guess. Normal parametric representations as general utilizes are straight forecast coefficient or smoothed representations and also different other representations are determined from these bit of the recogniser is to impact a aligning between groupings of talk vectors and the required principal picture strategies. Two issues male this extraordinarily extreme. Right off the bat, the mapping from pictures to talk isn't coordinated since diff erent fundamental pictures can enable ascent to similar talk sounds. Additionally, there are tremendous assortments in the made sense of it talk waveform because of the variability, personality, condition, and so forth. Besides the limits between pictures can't be perceived from the talk waveform. Subsequently, it isn't possible to regard the talk waveform as a gathering of connected dormant plans. The minute issue of not knowing the word limit zones can be maintained a strategic distance from by limiting the errand to disconnected word acknowledgment. As appeared in Figure3.2, this proposes the talk waveform analyzes to a solitary fundamental picture (example: word) taken from a settled dictionary. In spite of the reality that this less complex issue is to some degree artificial, it by the by has a large stretch out of viable function. Also, it fills in as an incredible preface for introducing the fundamental considerations of HMM-based acknowledgment some time recently managing with the more complex persistent discourse case. Thus, disconnected word acknowledgment utilizing HMMs will be overseen with first. Let each talked word be spoken to by an arrangement of discourse

vectors or recognitions O, characterized as takes after
$O = o_1,o_2,o_3\ldots\ldots,o_T$

where ot is the talk vector observed at time t. The separated word affirmation issue can by at that point be seen as that of figuring argmax I $\{P(wi|O)\}$ where w_i recommends i'th dictionary word. This probability isn't forms especially in any case utilizing Bayes' Run the appear gives

$p(w_i|O) = (p(O|w_i)p(w_i))/p(o)$

Subsequently, for a given course of action of before probabilities p(wi), the most possible talked word depends as they say on the probability p(O|wi). Given the dimensionality of the perception gathering O, the encourage estimation of the joint unforeseen likelihood p(o1,o2,o3,… … |wi) from frameworks of talked words are not practicable.. Back unwinding works well for condition when Well is ergodic, that means. there is move from any state to whatever other state. On the off chance that associated to a Gee of alternative building, this technique could allow an course of action that not to be a substantial way since a few moves are not embraced.

The Viterbi calculation picks the most excellent state gathering that adventures the likelihood of the state course of action for the given recognition progression[2]. Grant δt(i) an opportunity to be the maximal likelihood of state groupings of the length t that conclusion in state I and make the t in any case affirmations for the given model

$\delta t(i) = \max\{P(q(1), q(2),q(3),\ldots ....., q(t-1) ; o(1), o(2),o(3) ..... ,o(t) \mid q(t)= q_i).\}$

The viterbi calculation is a energetic programming calculation that utilizations an undefined design from the Forward calculation with the exemption of two contrasts:

1. It jobs maximation in put of summation at the recursion and end steps.

2. It screens the suppositions that enhance δt(i) for each

t and I, securing them in the N by T cross area ψ.

This cross section is utilized to recover the perfect state movement at the movement of backtracking.

Right when Initialization $\delta 1(i) = p_i\, b_i(o(1))$

$\psi 1(i) =$, I =1,2 ....., N

According to the over definition, ßT(i) does not exist. The recursion formal improvement as takes after.

$\delta t\ (\ j) = \max_i\ [\delta t - 1(i)\ a_{ij}]\ b_j\ (o9t))$

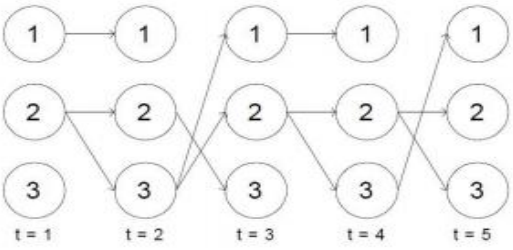
$\psi t(\ j) = \arg \max_i\ [\delta t - 1(i)\ a_{ij}]$

The last end is:

$P^* = \max_i\ [\delta T(i)]$

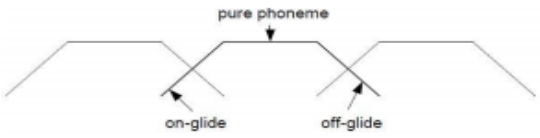
$q^*t = \arg \max i\ [\delta T(\ I\ )]$ Then the state grouping backtracking:
$q^*t = \psi t+1(\ q^*t+1)$, t = T - 1, T - 2 , T-3. . . . . , 1

calculation can be visualized utilizing a trellis as given bellow. The result appears of applying the Viterbi calculation to the Well to discover the state grouping comparing to O =10110.



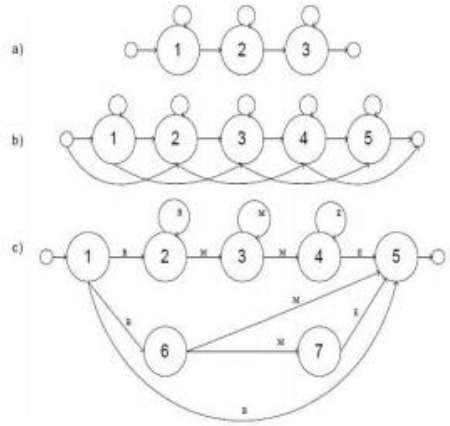
Figure(4): The Viterbi algorithm visualized by trellis

Well Topology for Speech Recognition Given an satisfactorily broad and operator planning set in light of the Viterbi calculation, the parameters of a Gee can be assessed utilizing topology of a Gee data. Over all else, choose what sort of unit in talk be talked to in Well. Various choices can be made, for occasion communicates, words, syllables, phonemes or other sub word units. Truth be told it is conceivable to utilize HMMs to appear any unit of talk, notwithstanding of the plausibility that the talk units are incapably passed on in Mumble. Words show up to be the most typical units to appear, since they are what we require to see and the dialect demonstrate moreover utilizes words as fundamental units. Semantically characterized sub word models utilize human specific data for isolating the parameter space. Acoustically characterized sub word units utilize.programmed calculations to examine the acoustic likenesses. Half and half models, utilizing both acoustic and phonetic learning moreover exist for Well based talk affirmation. Phone models are the most utilized sub word units. Since there are fair 40 to 50 phonemes in dialects like English and Dutch. HMMs in light of telephone models can be adequately prepared. Most topologies utilized as a portion of discourse acknowledgment depend on the doubt that there are three stages in the enunciation of a phone. In the fundamental organize the vocal tract is changing shape to verbalize the phone, this is known as the on-coast of the phone. In this organize there might be a few cover with the previous phone. In the moment arrange the sound of the phone is thought to be flawless and in the third arrange the sound is released and the vocal tract starts to travel to the taking after phone. This is known as the off-coast, a few cover with the taking after phone may happen here, the strategy is schematically illustrated as follows.



Figure(5): Three periods of Phonemes

The accompanying shows three model topologies that have effectively been utilized as a part of different discourse recognizers. The main model (figure 6.a) is a basic three state left-right model with state subordinate yield probabilities. The first and last littler circles in the figure speak to section and leave expresses, these are alleged invalid states, they don't produce perceptions and are just used to connect the models. The second model (figure 6.b) has five states, yet gives moves that avoid the succeeding state; in this way it is conceivable to go through just three stages. This model likewise has state subordinate yield probabilities. The last model (figure 6.c), which is the model utilized by IBM, has seven states and twelve moves with move subordinate yield probabilities.Three gatherings of yield probabilities are tied, comparing with the three stages in a phoneme. In the figure the start stage (on-skim) is set apart with B the center stage with M and the end stage (off float) with E. As an outcome this model just has three distinctive likelihood circulation capacitie.



Figure(6): Show topologies for phoneme Units

Fitting the Phonemes together The past portions talked around tongue models and acoustic models however once phonemes level are framed, in HMMs how do the phonemes are assembled? we require to take note that a lingo model can be seen as a framework of states (the words) associated by moves with probabilities appended to them. At the conclusion of the day a tongue show can be seen as a Markov Show. Misusing the way that introducing HMMs into a Well prompts another Well we can supplant the word states by the comparing word or phone level HMMs bringing approximately one immense Well. On the off chance that there ought to be an event of phone level HMMs[10], the phone models must be connected to shape a word, before substituting. Figure.7 illustrates a chart dialect show for an course of action of the words one and two with its pledge models instantiated.

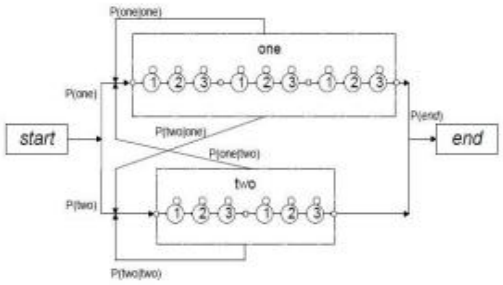


Figure7: Composite HMM for Viterbi recognition

4. Results and Discussions

HMM-based speech to text conversion system, five audio files like banana, computer and apple are modelled in HMM(Hidden markov model). The original signal at the sampling rate of 8 kHz are explained in the Figure.8.

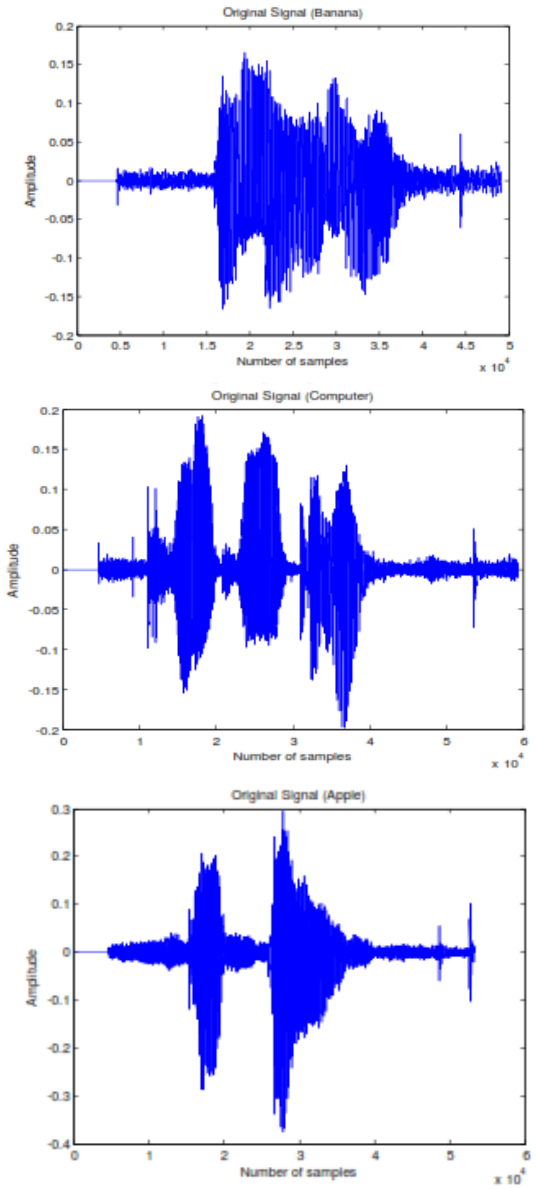


Figure 8.Nuber of samples versus amplitude of three original signals.

5. Conclusion

In this we found that separation measures between the Hidden Markov Models (HMMs) are utilized for the investigation and comprehension of Non Audible Murmur(NAM) Speech. Around then when Compared with the those of typical speech has been demonstrates that attributable to the diminished spectral space of NAM discourse and furthermore the HMM separations are additionally decreased. Loss of lip radiation and furthermore body transmission are go about as a low pass channel in NAM receiver. Because of this the outcome is the segments of a recurrence are lessened in a NAM flag . Using this Model we can design a speech detector ,in such a way that the detector will be used in Army by the Commander to give the messages to his soldiers safely without any misleads.

References

[1] BDenby, TSchultz, KHonda, THueber, JMGilbert, and JS.Brumberg, “Silent speech interfaces,” Speech Communication., vol52, no4,pp270–287, 2010.

[2] T.Rajeshkumar and G.R.Suresh “Examination of militants utiliz-ing NAM microphone and wireless handset for murmured speech in a view of concealed markov model.” IIRJET In 2017.

[3] S.-CJou, TSchultz, and AWaibel, “Adaptation for soft whisper recognition using a throat microphone,” in ProcINTER-SPEECH, Jeju Island, Korea, Sep2004, pp1493–1496.

[4] TSchultz and MWand, “ Modeling coarticulation in EMG-based continuous speech recognition,” Speech Communication., vol52, no4, pp.341–353, 2010.

[5] THueber, E.-LBenaroya, GChollet, BDenby, GDreyfus, and M.Stone, “Development of a silent speech interface driven by ultra-sound and optical images of the tongue and lips,” Speech Commu-nication., vol52, no4, pp288–300, 2010.

[6] ASubramanya, ZZhang, ZLiu, and AAcero,“Multisensory pro-cessing for speech enhancement and magnitude-normalized spectra for Speech modeling,”SpeechCommun.,vol.50,no.3,pp.228–243,2008.

[7] Y.Nakajima,H.Kashioka,N.Cambell,andK.Shika no.”Non-audible murmur (NAM) recognition,” IEICE TransInfSyst., volE89-d, no1, pp1-8,2006

[8] Panikos Heracleous,Tomomi Kaino,Hiroshi Saruwatari and Kiyohi-ro Shika-no”Unvoiced speech recognition using Tissue conductive acoustic sensor” In 2006

[9] T.Toda,K..Nakamura,T.Nagai,T.Kanio,Y.Nakajima, and KShi-kano, “Technologies for processing body-conducted speech detect-ed with non-audible murmur microphone,” in proc.INTERSPEECH, Brighton, U.K., Sep2009, pp,632-635.

[10] Yamato Ohtani and Tomkoi “non parallel preparing for nu-merous to numerous Eigen voice change” IEEEIn 2010.