ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Voice Controlled Network with Enhance Security

Radha Sharma, Dr. Dinesh Kumar

M.Tech Student, Department of Computer Science and Engineering, SRCEM at Palwal, MD University, Haryana, India.

Professor & HOD, Department of Computer Science and Engineering, SRCEM at Palwal, MD University, Haryana, India.

ABSTRACT: Imagine the scenario, you've simply left home or workplace and you've forgotten to show your computer and network off, your native electricity company are Terpsichore with joy and your colleague may come back to your table to visualize that employment you've been viewing or functioning on. What to do? does one circle and go all the means back? Well, currently there's no want, as a result of you'll be able to truly close up your computer/network (and a full ton more) by simply clicking on the sensible phone. There's no price concerned altogether quite these items. we have a tendency to solely want Associate in Nursing planned application to close up your machine, we have a tendency to square measure only 1 click away to motility the pc down or execute another operations like taking part in music, accessing files, obtaining still screen shots, stopping the various applications that square measure running on the pc or network.

I.INTRODUCTION

Earlier users accustomed head to every and each individual machine within the network and access resources on that. There no thanks to employing a single server that is connected in the same native space network. There have been no ways in which to impose these rules for a distant server.

There is additionally the windows utility known as remote desktop which supplies North American country the power to remotely hook up with a computer/PC within the network, when obtaining connected to the pc the screen of the pc seems on the machine from wherever folks are connecting. when the affiliation is successful folks will management the laptop as if it's their own laptop and folks are dominant it with their keyboard and mouse. however this windows utility is totally completely different from my project because it don't connect the machine however still will management the resources to lock and unlock them. this fashion is saves the process power of each the server and therefore the shopper computers, therefore dashing up the method and additionally provides mobility via a hand-held phone.

In the projected resolution it'll write associate degree application in java with to completely different parts as server element and shopper system with increased security. The shopper is largely the mobile application that has been created victimization Java & humanoid is put in our transportable and therefore the server is during a machine or laptops. It communicates among them through the WiFi affiliation and permits the remote controller of the pc to manage the machines. within the remote laptop associate degree application written in .java works within the background that executes the management commands.. A graphical user interface which is very user friendly and very easy to learn and understand for the end users is also developed. Speech recognition incorporates a long history, minimally chemical analysis back to the 1952 Bell Labs paper describing a way for digit recognition. Similarly, machine learning incorporates a long history, with vital development within the branch normally referred to as neural networks conjointly going back to a minimum of the Fifties. Speech recognition ways converged by 1990 into applied math approaches supported the hidden mark-off model (HMM), whereas artificial neural network (ANN) approaches in common use cared-for converge to the multilayer perceptions (MLP) incorporating back-propagation learning. a lot of recently, there has been significant interest in neural network approaches to phone recognition that incorporate several of the rhetorical characteristics of MLPs (multiple layers of units incorporating nonlinearities), however that don't seem to be restricted to back propagation for the training technique. It ought to be noted en passant that the earliest ANN coaching ways didn't use error back propagation; as an example, the Discriminated Analysis reiterative style (DAID) technique

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

developed at Cornell within the late Fifties incorporated multiple layers that were severally trained, victimization mathematician kernels at the hidden layer and a gradient coaching technique at the output layer. it's been necessary for machine intelligence researchers to conduct experiments with modest-sized tasks so as to allow intensive explorations; on the opposite hand, conclusions from such experiments should be drawn with care, since they usually don't scale to larger issues. For this reason, among others, several researchers have gravitated towards large-scale issues like large-vocabulary speech recognition, which regularly incorporate many voluminous input patterns, and might need the coaching of tens of voluminous learned parameters [4]. Given the maturity of the speech recognition field, competitive performance usually needs the utilization of difficult systems, that any novel part plays a persona. trendy speech recognition systems, as an example, incorporate massive language models that use previous info to powerfully weight hypothesized utterances towards task-specific expectations of what could be same. Thus, it are often troublesome to visualize the advantage of a brand new technique. However, if up speech recognition is our goal, there's no alternative however to look at an outsized scale task, though smaller tasks are often accustomed validate code and rule out obvious issues with a plan. On the opposite hand, tiny tasks may also be each realistic and difficult; as an example, a little vocabulary recognition task that comes with fluently spoken words during a moderate quantity of noise and/or reverberation will yield vital insight concerning the lustiness of a projected technique. This paper can describe a number of the ways developed over the last decade that incorporate multiple layers of computation to either offer massive gains for crying speech on small-vocabulary tasks or modest however vital gains for high-SNR speech on large-vocabulary tasks. In every case the stress are to explain ways that have exploited structures incorporating each {a massive an outsized an oversized} range of layers (the depth) and multiple streams victimization MLPs with large hidden layers (the width). In some cases the underlying model is a minimum of an initio generative (as with the utmost chance coaching employed in standard ASR systems before discriminative training), however in alternative cases the ways are discriminative from the beginning. the main focus here are on what are currently classical ways, furthermore as newer approaches creating use of discriminatively trained options. In most cases these systems are inherently heterogeneous, incorporating a sequence of machine layers that perform differing functions. the category of systems incorporating deep belief networks, that are essentially generative nature however conjointly solid in their kind and coaching, are emphasized during this paper.

II. LITERATURE SURVEY

As of this writing, progressive automatic speech recognition (ASR) systems incorporate quite few layers of process before the output of word sequences. the method starts with many layers of signal process (e.g., windowing, short-run spectral analysis, vital band spectral integration and campestral transformation). it's true that these stages area unit generally enforced with fastened parameters; on the opposite hand, there has been recent work that has shown enhancements exploitation learned parameters for a nonlinear perform of the spectral values, impressed by the amplitude compression that's evident in human hearing [5]. However, even for different systems, it's quite common to rework the spectrum by compression or increasing it during a method known as vocal tract length normalization [6]. Despite its name, VTLN doesn't need any live of the vocal tract, however uses applied math learning techniques to work out the maximum-likelihood compression/expansion of the spectrum for every clustered vocalization or speaker (often derived from associate degree unsupervised learning algorithm); these approaches area unit supported associate degree underlying generative model. Another common element is Linear Discriminate Analysis (LDA) or its less forced full cousin, Heteroscedastic Linear Discriminant Analysis (HLDA), every of that is trained to maximize phonetic discrimination. This layer transforms campestral options, generally over many past and future acoustic frames, into a replacement observation sequence for the popularity system. The ensuing options area unit then accustomed train an oversized range of Gaussians that area unit utilized in combination to come up with likelihoods for specific speech sounds in context. Note that each the individual Gaussians and their mixture coefficients area unit trained, which Expectation–Maximization is employed for coaching since the weight of every element within the mixture is unknown a priori, even in coaching. Following coaching with a maximum likelihood criterion, objective functions like most mutual info (MMI) or minimum phone error (MPE) [7] area unit generally accustomed train the mathematician parameters discriminatively. The parameters of this acoustic model area unit then altered more for testing by incorporating one among many connected strategies for adaptation, as an example Maximum-Likelihood regression (MLLR) [8]. the whole acoustic chance estimation scheme is then utilized in combination with a language model likelihood estimation, that has been trained during a supervised fashion on an oversized range of words; in addition, there area unit sometimes multiple sources of word prediction info (such as massive quantities of written language and

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

smaller amounts of transcribed spoken words) so weight and back off parameters should be learned. The interpolation coefficients between language and acoustic level log likelihoods are learned, as area unit varied different recognizer-specific parameters. Finally, the most effective recognizers generally incorporate multiple complete systems that mix their info at varied levels, like what's known as cross-adaptation, during which coaching targets for one system comes from the opposite [9].

All of the on top of assumed one stream of speech coming into the system. However, it's changing into additional common in observe to possess a minimum of to speech signal streams, one from every of to or additional microphones. Combination techniques unremarkably incorporate unsupervised learning strategies to see the simplest combination of the electro-acoustic transducer outputs [10]. this can be far away from an entire list of common ASR elements; however it ought to do to indicate the reader that even ASR systems that don't habitually incorporate artificial neural networks or alternative expressly stratified machine learning mechanisms area unit each deep (many layers of computation) and wide (many totally different components combined). These layers typically have fastened parameters, however in several cases they're learned, and infrequently with Associate in nursing underlying generative model. The Section III can review a category of extra learned layers that are a dscititious in some systems to nonlinearly method the options that area unit fed to the applied math engine.

In 2000, as part of a European Telecommunications Standards Institute (ETSI) competition for a new Distributed Speech Recognition standard [11], an approach to speech recognition was developed that was called Tandem [12]. Drawing on MLP techniques developed in the context of computing discriminant emission probabilities for HMMs [13], this approach generated features for the HMM that were trained for phonetic discrimination. As in the earlier techniques, the MLP in the newer approach is trained with phone label targets, so that it estimates state or phone posterior probabilities; outputs from multiple MLPs are sometimes combined to improve the probability estimates. The typical initial system used a single nonlinear hidden layer. However, later architectures incorporated more layers; for instance the so-called TRAPS system used such an MLP for a half second of the time sequence of energies for each critical band of the spectrum (or for each set of three bands, with overlap) [14], followed by a combination component that comprised an additional MLP with its own hidden layer. This was learned separately, so that there was no attempt to back propagate errors all the way back through the system. A later form of this system called HATs [15] was trained by taking the input-to- hidden nonlinear transformations from each critical band and using their outputs to feed the final combination MLP (Fig. 1). In a number of large tasks (American English conversational telephone speech, American English broadcast news, Mandarin broadcast news, Arabic broadcast news), a further combination of HATs output and an MLP processed version of standard PLP features was used to provide significant (roughly 10% relative) reductions in errors. This was one of the largest reductions shown from any improvement in the systems under test [9].

MORGAN: DEEP AND WIDE: MULTIPLE LAYERS IN ASR

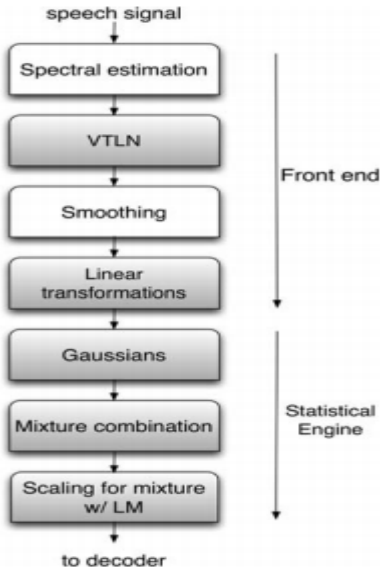

Fig. 1. Computational layers for standard single stream large vocabulary speech recognition acoustic model.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Shaded boxes represent layers with at least some learned parameters. VTLN stands for Vocal Tract Length Normalization (defined in text). Not shown are the decision trees that determine the structure of the models in the statistical engine; these also have learned parameters.

A. Other Approaches

At IBM, researchers developed a technique called Featurebased Minimum Phone Error (fMPE) [16], which incorporated the MPE error criterion at the feature level. In this approach, the features were generated by training a large number of Gaussians over the acoustic sequence2 and computing temporally local posteriors. In practice it provided similar improvements to either MPE training of the acoustic models or to the MLP-based approach described above. Combinations of these methods have also been explored in [17]. Other approaches have been built on a hierarchical feature approach, for instance training Tandem features for high temporal modulation frequencies and using them, appended to low temporal modulation frequencies as input for a second network generating Tandem features (Fig. 2). Thus, one path through the networks encountered four layers of processing by trained parameters while the other encountered two. This method provided significant improvement on a difficult ASR task requiring recognition of speech from meetings [18], and later was demonstrated to provide significant improvements in character error rate for a large-vocabulary Mandarin broadcast news evaluation [19].

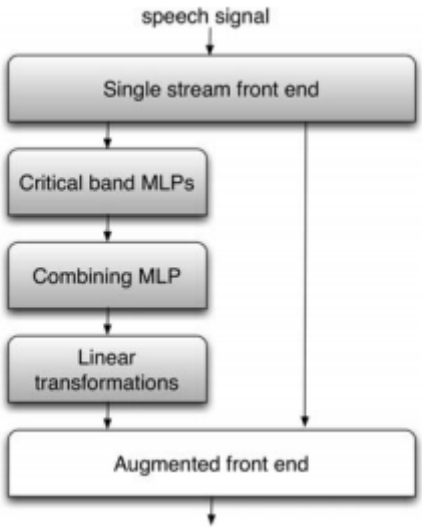

Fig. 2. Computational layers for the TRAPS or HATS version of Tandem processing, where the first layer summarizes the front end processing from Fig. 1. All layers except the last have learned parameters. Each of the MLP layers has nonlinear hidden units. The critical band MLPs can either be trained separately from the combining MLP (as in HATs) or in one large training with connectivity constraints (as in Chen's Tonotopic MLP). The simplified figure does not show that the input to the MLP stage is from the pre-smoothed spectral values, while the standard components of the feature vector are cepstral values with other transformations (e.g., derivatives, HLDA, etc.) All MLPs were trained with phone targets, and generated estimates of phone posteriors. Each output from the MLP is either taken prior to the final nonlinearity or else after computing the log probability. The linear transformation in this and later figures typically consists of principal component analysis (PCA), which requires unsupervised learning to determine the orthogonal dimensions with the greatest variance.

III. PROPOSED SYSTEM

So this was all about training phase of language models. The outputs of this phase are three probability files for each language as described above. Now before discussing about other phases lets discuss briefly how these are going to use these probability files. Let's suppose "a b c d e f "is a phoneme sequence, where a, b, c, d, e and f are different phonemes, came for recognition during testing phase, then different probabilities related to it will be using Distance Minimum techniques over below mentioned aspects along-with below mentioned security algorithm :-

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Unigram_Prob = $P(a|X)* P(b|X) *P(c|X)*P(d|X)*P(e|X)*P(f|X)$……..(1)
Bigram_Prob = $P(b|X, a)* P(c|X, b) *P(d|X, c)*P(e|X, d)*P(f|X, e)$……...(2)
Trigram_Prob = $P(c|X, a, b)* P(d|X, b, c) *P(e|X, c, d)*P(f|X, e, f)$…….(3)

SECURITY ALGORITHM

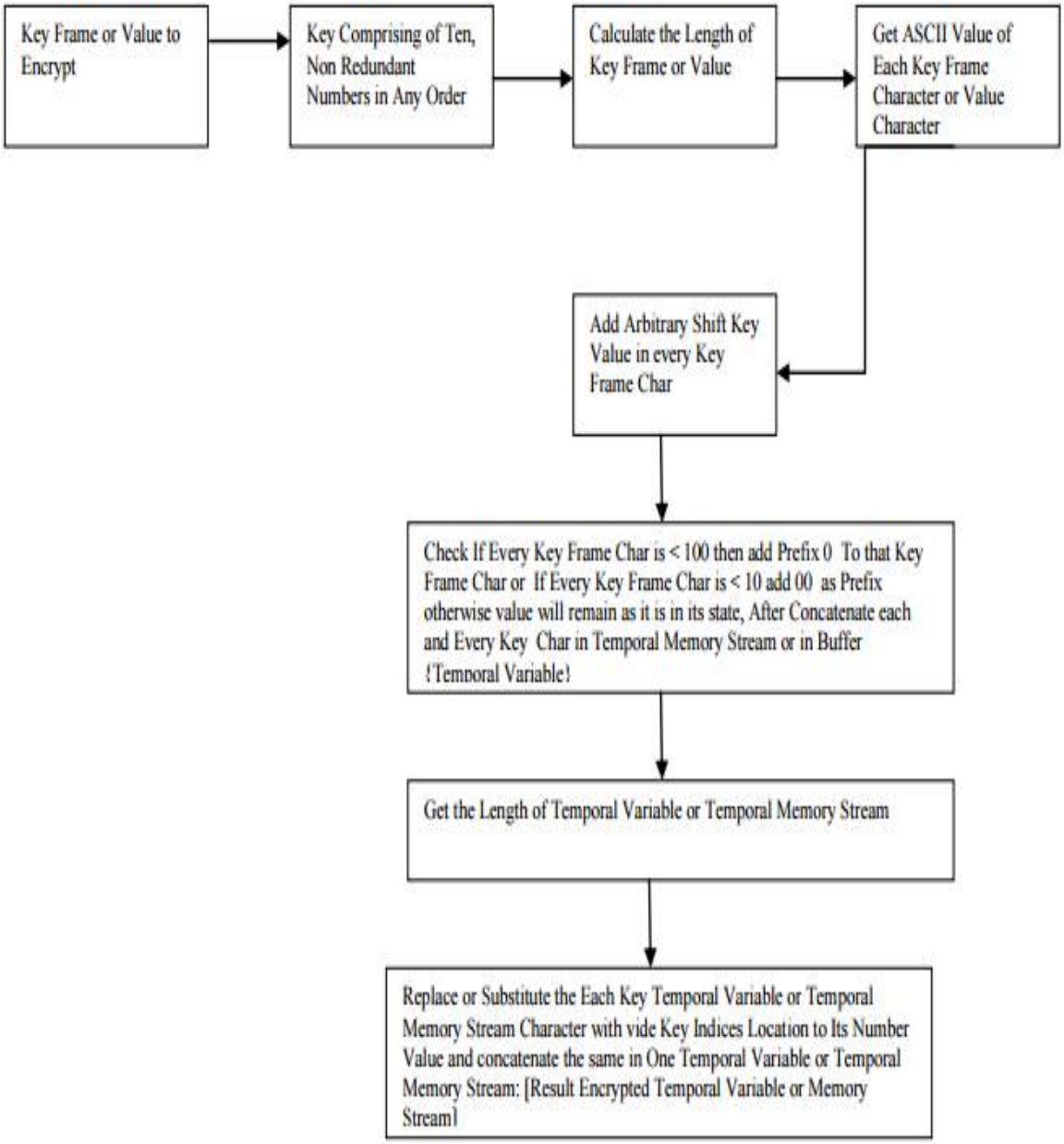

Figure3.1 Flow Graph of encryption algorithm

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

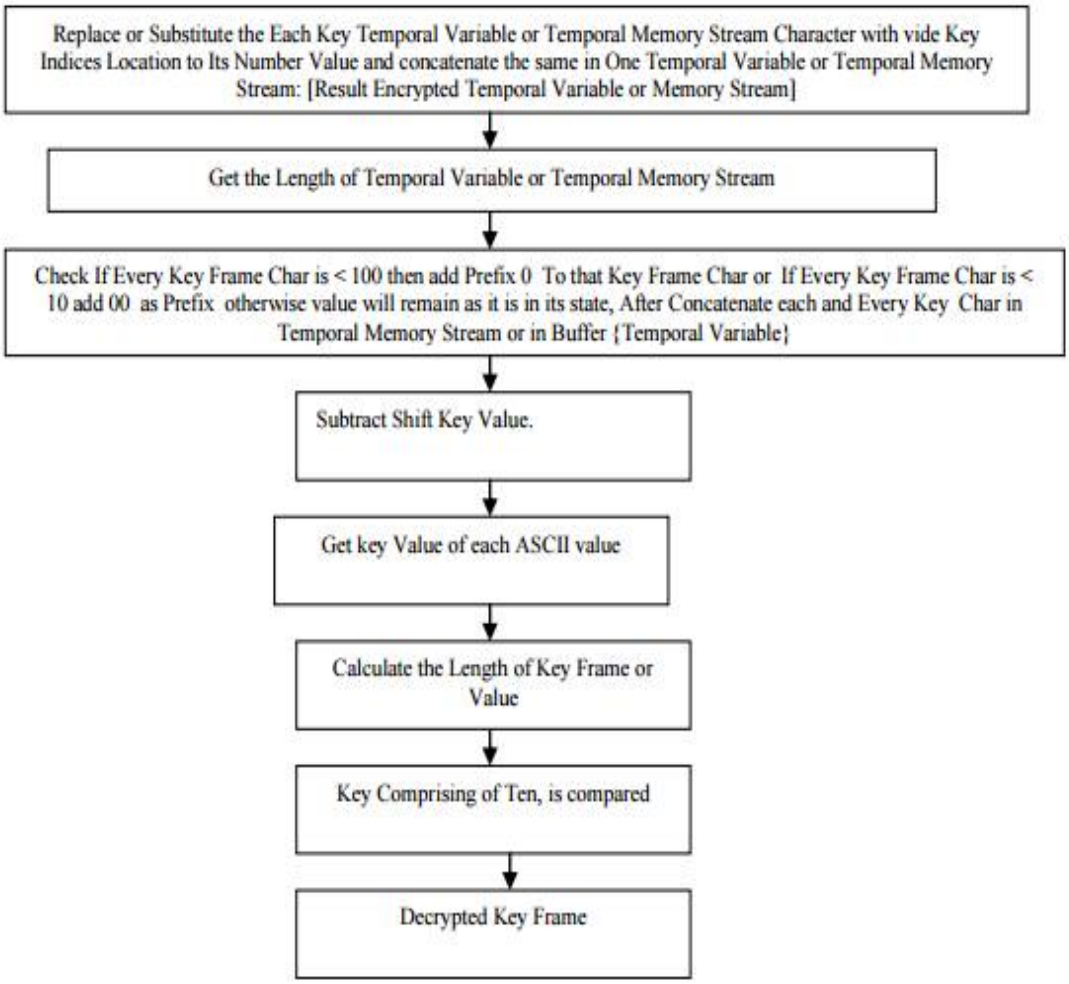

Figure3.2 Flow chart of Decryption

IV. RESULTS

In the speech recognition phase, four different (2 male and 2 female) speakers are asked to speak the same word ten times from the given list of words. The speakers are then asked to utter the same words in a random order and the recognition results noted. The percentage recognition of a speaker for these words is given in the table 1 and efficiency chart is shown in figure 5 for the same. The overall efficiency of speaker identification system is 95%. In speech recognition phase, the experiment is repeated ten times for each of the above words. The resulting efficiency percentage and its corresponding efficiency chart are shown in table 1. The overall efficiency of a speech recognition system obtained is 98%.

The below table 1. depicts the results achieved from the punctuation pronounced using Trigram, Unigram and Bigram analogies with trained network pertaining the respected vocabulary and patterns to measure the efficiency and accuracy with different voice patterns and jaws expressions over the air using the mobile device and results the below table with respective percentage and collective average percentage.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Table 1. Speaker identification results

Words	Female Speaker 1	Female Speaker 2	Male Speaker 3	Male Speaker 4
Computer	90%	100%	100%	90%
Read	100%	100%	100%	100%
Mobile	90%	100%	90%	90%
Man	100%	70%	100%	100%
Robo	80%	100%	100%	100%
Average %	92%	94%	98%	96%

T

V. CONCLUSION

In the speaker identification phase, MFCC and Distance Minimum techniques have been used over Unigram, Bigram, and Trigram analogy . These techniques provided more efficient speaker identification system. The speech recognition phase uses the most efficient HMM Algorithm. It is found that Speaker recognition module improves the efficiency of speech recognition scores. The coding of all the techniques mentioned above has been done using Java. It has been found that the combination of MFCC and Distance Minimum algorithm gives the best performance and also accurate results in most of the cases with an overall efficiency of 95%. The study also reveals that the HMM algorithm is able to identify the most commonly used isolated word. As a result of this, speech recognition system achieves 98% efficiency.

REFERENCES

[1] K. H. Davis, R. Biddulph, and S. Balashek, "Automatic recognition of spoken digits," J. Acoust. Soc. Amer., vol. 24, no. 6, pp. 627–642, 1952
[2] G. Dahl, M. Ranzato, A. Mohamed, and G. E. Hinton, "Phone recognition with the mean-covariance restricted Boltzmann machine," in Advances in Neural Information Processing 23. Cambridge, MA: MIT Press, 2010.
[3] S. S. Viglione, "Applications of pattern recognition technology," in Adaptive Learning and Pattern Recognition, J. M. Mendel and K. S. Fu, Eds. New York: Academic, 1970, pp. 115–161.
[4] D. Ellis and N. Morgan, "Size matters: An empirical study of neural network training for large vocabulary continuous speech recognition," in Proc. ICASPP, 1999, pp. 1013–1016.
[5] Y.-H. Chiu, B. Raj, and R. Stern, "Learning based auditory encoding for robust speech recognition," in Proc. ICASSP, 2010, pp. 428–4281.
[6] J. Cohen, T. Kamm, and A. Andreou, "Vocal tract normalization in speech recognition: compensation for system systematic speaker variability," J. Acoust. Soc. Amer., vol. 97, no. 5, pt. 2, pp. 3246–3247, 1995.
[7] D. Povey, "Discriminative training for large vocabulary speech recognition," Ph.D. dissertation, Cambridge Univ., Cambridge, U.K., 2004
[8] C. J. Leggetter and P. C. Woodland, "Maximum likelihood linear regression for speaker adaptation of continuous density HMMs," Speech Commun., vol. 9, pp. 171–186, 1995.
[9] A. Stolcke, B. Chen, H. Franco, V. R. R. Gadde, M. Graciarena, M.-Y. Hwang, K. Kirchhoff, A. Mandal, N. Morgan, X. Lei, T. Ng, M. Ostendorf, K. Sonmez, A. Venkataraman, D. Vergyri, W. Wang, J. Zheng, and Q. Zhu, "Recent innovations in speech-to-Text transcription at SRIICSI-UW," IEEE Trans. Audio, Speech, Lang. Process., vol. 14, no. 5, pp. 1729–1744, Sep. 2006.
[10] M. Seltzer, B. Raj, and R. Stern, "Likelihood maximizing beamforming for robust hands-free speech recognition," IEEE Trans. Speech Audio Process., vol. 12, no. 5, pp. 489–498, Sep. 2004.
[11] H.-G. Hirsch and D. Pearce, "The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in Proc. ISCA ITRW ASR2000 Autom. Speech Recognition: Challenges for the Next Millennium, Paris, France, Sep. 2000.
[12] H. Hermansky, D. Ellis, and S. Sharma, "Tandem connectionist feature extraction for conventional HMM systems," in Proc. ICASSP, Istanbul, Turkey, Jun. 2000, pp. 1635–1638.
[13] H. Bourlard and N. Morgan, Connectionist Speech Recognition: A Hybrid Approach. Norwell, MA: Kluwer, 1993.
[14] H. Hermansky and S. Sharma, "TRAPS—Classifiers of temporal patterns," in Proc. 5th Int. Conf. Spoken Lang. Process. (ICSLP'98), 1998, pp. 1003–1006.
[15] B. Y. Chen, "Learning discriminant narrow-band temporal patters for automatic recognition of conversational telephone speech," Ph.D. dissertation, Univ. of California, Berkeley, 2005.
[16] D. Povey, B. Kingsbury, L. Mangu, G. Saon, H. Soltau, and G. Zweig, "fPME: Discriminatively trained features for speech recognition," in Proc. IEEE ICASSP'05, 2005, pp. 961–964.
[17] J. Zheng, O. Cetin, M.-Y. Hwang, X. Lei, A. Stolcke, and N. Morgan, "Combining discriminative feature, transform, and model training for large vocabulary speech recognition," in Proc. IEEE ICASSP'07, Honolulu, HI, Apr. 2007, pp. 633–636.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

[18] H. Hermansky and F. Valente, "Hierarchical and parallel processing of modulation spectrum for ASR applications," in Proc. IEEE ICASSP'08, 2008, pp. 4165–4168.
[19] F. Valente, M. Magamai-Doss, C. Plahl, and S. Ravuri, "Hierarchical processing of the modulation spectrum for GALE Mandarin LVCSR system," in Proc. Interspeech'09, Brighton, U.K., 2009.
[20] N. Morgan, Q. Zhu, A. Stolcke, K. Sonmez, S. Sivadas, T. Shinozaki, M. Ostendorf, P. Jain, H. Hermansky, D. Ellis, G. Doddington, B. Chen, O. Cetin, H. Bourlard, and M. Athineos, "Pushing the envelope—Aside," IEEE Signal Process. Mag., vol. 22, no. 5, pp. 81–88, Sep. 2005.
[21] J. P. Pinto, "Multilayer perceptron based hierarchical acoustic modeling for automatic speech recognition," Ph.D. dissertation, EPFL, Lausanne, Switzerland, 2010.
[22] S. Zhao, S. Ravuri, and N. Morgan, "Multi-stream to many-stream: Using spectro-temporal features for ASR," in Proc. Interspeech, Brighton, UK, 2009, pp. 2951–2954.
[23] G. Hinton, personal communication. 2010.
[24] L. Chase, "Error-responsive feedback mechanisms for speech recognizers," Ph.D. dissertation, Robotics Inst., Carnegie Mellon Univ., Pittsburgh, PA, 1997.
[25] S. Wegmann and L. Gillick, Why has (reasonably accurate) automatic speech recognition been so hard to achieve? Nuance Commun., 2009.
[26] J. M. Baker, L. Deng, J. Glass, S. Khudanpur, C. Lee, N. Morgan, and D. O'Shaughnessy, "Research developments and directions in speech recognition and understanding, part 1," IEEE Signal Process. Mag., vol. 26, no. 3, pp. 75–80, May 2009.
[27] J. M. Baker, L. Deng, J. Glass, S. Khudanpur, C. Lee, N. Morgan, and D. O'Shaughnessy, "Research developments and directions in speech recognition and understanding, part 2," IEEE Signal Process. Mag., vol. 26, no. 4, pp. 78–85, Jul. 2009.