Articulation Error Detection Techniques and Tools: A Review

Khushbu Bansal
Department of Computer Science and Engineering
PEC University of Technology,
Chandigarh, India

Shailendra Singh
Department of Computer Science and Engineering
PEC University of Technology,
Chandigarh, India

Dharam Vir
Department of Otolaryngology
PGIMER, Chandigarh, India

Swati Sharma
Department of Computer
Science and Engineering
PEC University of Technology,
Chandigarh, India

ABSTRACT
Speech is the major source of communication. Articulation errors affect a person's speech in adverse way. Speech language pathologists have to calculate the articulation errors manually amongst the persons suffering from speech problems. This task is very time consuming and exhaustive. Therefore, a system needs to automate this task. This paper presents all the advancements done in the field of speech recognition right from speech classification, feature extraction, speech models and tools by which an articulation error detection system can be built. The objective of this paper is to compare various methods by which an efficient articulation error detection system can be formulated**.**

Keywords
Speech recognition, articulation errors, picture naming task, feature extraction, hidden markov model, vector quantization.

1. INTRODUCTION
Automatic Speech recognition (ASR) is a procedure by which a computer identifies the spoken words. Speech Recognizer is a tool which understands the spoken word and act afterwards so that, it can be used to find speech disorders. Speech disorder is a type of communication disorder where normal speech is disrupted. This is mainly categorized into two forms Articulation errors and phonological errors. Articulation errors refer to a person's inability to produce certain phoneme (alphabet or a word) correctly. While phonological disorders refer to person disability of learning the pronunciation of a certain word. This disorder is mainly related with children but if it is not corrected at right time, it may persist until adulthood called residual errors [1]. According to speech pathologists a normal child (without any articulation disorder) is able to produce certain alphabets in certain year or months. For instance, a child between 1-3 years of age is able to pronounce p, m, h, v alphabets. At the age of 3-4 years children are able to utter n, b, k, g, and t correctly. At the age of 4-5 they are able to speak, d, f, ch, sh, z and at the age of 8-9 they are able to pronounce r, s, z. If these letters are not articulated by children correctly, it means they have speech problems. Speech problems can also take place due to physical inability or mental disability of a person

such as children suffering from hearing disability, cerebral palsy, mental retardation and neonatal jaundice. Further sections contain the insight about types of ASR, applications, problems in ASR, feature extraction methods and various speech recognition models.

2. CLASSIFICATION OF SPEECH RECOGNITION SYSTEMS
Speech recognition systems can be classified as following

2.1 Forms of Utterances [2]
2.1.1 Isolated words
It requires isolated word or single word at a time. It has two phases Listen or Non-Listen during which it ask speakers to gap between two consecutive words.

2.1.2 Connected Words
These have similarity to isolated ones except that there is lesser pause in between words.

2.1.3 Continuous Speech
It allows speaker to talk normally while computer recognize the content.

2.1.4 Spontaneous Speech
Spontaneous speech is similar to continuous one. It is a natural sound and not a rehearsed one. The system built should be trained to recognize spontaneous words.

2.1.5 Voice Verification/Identification
The system is able to identify the person speaking. It requires regressive training and testing.

2.2 Types of speaker model
2.2.1 Speaker Dependent Models
These are developed for known speakers.This model is mostly correct for known speakers and less correct for unknown ones. These are easy to develop but not as flexible as other models.

2.2.2 Speaker Independent Models
These are designed for variety of speakers as these are not speaker dependent. These are comparatively difficult to develop but more flexible.

2.2.3 Speaker Adaptive
This kind of model adapts its characteristics according to the new speaker. It lies between dependent and independent models.

2.3 Types of Vocabulary

The size of vocabulary can be one of the following

2.3.1 Small

It contains small (10's) words.

2.3.2 Medium

It contains (100's) medium words.

2.3.3 Large

It contains (1000's) large words.

2.3.4 Very large

It contains almost ten thousand words.
The size decides how complex and accurate a system is [3].

3. DEVELOPMENT OF ASR SYSTEMS

Although speech recognition has become very famous in last decades and it has been used by millions of people. The focus is now on ASR systems that detects articulation error patterns. Before analyzing how to make those systems, a brief history in the progress of ASR system must be known. The major developments in ASR are given in Table1 [3] [8].

Table 1. Development Of Asr System

Year	Progress
2000	Variation Bayesian Technique
2002	Large vocabulary Systems & Controlled Environment
2004	Unlimited vocabulary
2007	Speech Broadcasting through Machines
2009	Vocabulary (large), Arbitrary Environment
2010	Synthesized Speech
2012	Multilingual system for speech enabled devices
2013	Real time speech recognition
Future direction	Articulation error detection in multilingual languages by speech recognition process

4. BASIC MODEL OF SPEECH RECOGNITION

It is very necessary to understand the basic model of speech recognition before digging deeper. Figure1 shows the basic model and its mathematical representation. It contains acoustic front end, acoustic model unit and search unit shown in figure1 [5].

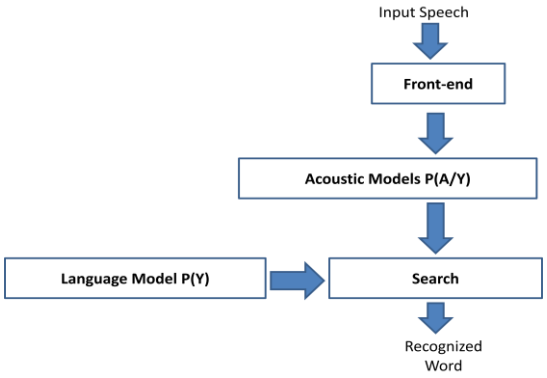

Fig 1: Model of speech recognition

It uses an approach with sequence of words denoted as X, an observation (acoustic) as Y, thus providing probability as P(X,Y). The aim is to find the word string which is referred as recognized utterance. ASR takes an acoustic waveform in the form of input and produces string of words as output. When the input is in acoustic form e.g. $X= X_1, X_2.....X_n$, the output given by the word sequence would be $Y = Y_1, Y_2.....Y_m$ with probability of P (Y|X). It can also be expressed by Bay's Theorem as given below (1) [3].

$$P(Y|A) = P(A|Y)\,P(Y) / P(A) \qquad (1)$$

where, A is acoustic signal, Y words spoken, P (A|Y) is the acoustic model and P(Y) is language model.

Language model predicts a small set of words based on knowledge of a finite number of previous words. Search is crucial to the system as it finds the most appropriate word sequence.

5. PHASES OF ASR

ASR consists of following two phases

5.1 Training phase

In training phase the speech is preprocessed and various feature extraction techniques along with modeling takes place. These will be discussed further in detail.

5.2 Recognition Phase

In this phase, acoustic analysis of speech signal and conversion of feature vectors using a specific algorithm takes place, to generate results according to how we want our output (e.g. generation of articulation error word). ASR has only the ability to recognize the word for which it is trained. This is the main phase while building an ASR.

5.3 Testing Phase

Testing phase includes testing the system in real time. It decides whether the system is feasible to use or not. So, it is one of the most important phases.

6. ISSUES WITH ASR

Difficulties may arise while building the speech recognition system. Table2 defines the difficulties or the steps to be taken while building an ASR. The difficulties may vary according to what kind of ASR is being built [4] [5].

Table 2. Issues With Asr System

Problems	Description
Transducer	Microphone/Headphone/Telephone
Channel	Amplitude/Distortion/Echo
Environment	Noisy, Quiet
Style of Speech	TONE-normal/shout, Production of words Isolated/continuous/spontaneous
Vocabulary	Specific/General
Body Language	Moving(wavy)/Stationary
Dialects	Regional/Social
Speaker Sex	Male/Female

7. APPLICATIONS OF ASR

Table 3 shows insight of the areas where ASR is used most often. Domain, input pattern and pattern classes are the things that needs to be noted [5]. It is also observed that ASR can be used in almost all fields that come across in day to day lives

Table 3. Applications Of Asr

Domain	Application	Input Pattern	Pattern Class
Speech/ Communication	Telephone Directory	Speech Waveform	Spoken Words
Education	Teaching foreign language students to speak a word correctly	Speech Waveform	Spoken Words
Physically Handicapped	Enabling students to enter Text verbally	Speech Waveform	Spoken Words
Electronics	Computers, Video Games	Speech Waveform	Spoken Words
Domestic	Refrigerators, Washing Machine, Oven	Speech Waveform	Spoken Words
Artificial Intelligence	Robotics	Speech Waveform	Spoken Words
General applications	Air Traffic Control, Call Home	Speech Waveform	Spoken Words
Medical Field	Medical Transcriptions	Speech Waveform	Spoken Words
Military	Telephony and people with disabilities	Speech Waveform	Spoken Words

8. METHODS

Methodology is the main prospective while building an ASR. On every stage the methods used, should be such that the system built as a whole gives maximum accuracy. To detect articulation errors knowledge about errors should be there. There are mainly four types of articulation errors.

- *Omissions*

An alphabet is omitted by the person. 'PAY' is spoken instead of 'PLAY' by the person.

- *Additions*

Additions of extra sound e.g. 'DOGUH' is uttered instead of 'DOG', 'BUHLACK' instead of 'BLACK' etc.

- *Substitution*

Substitution of an alphabet e.g. 'DUT' is uttered in place 'DUCK', 'THOAP' instead of 'SOAP' and 'WED' instead of 'RED'.

- *Distortion*

Distortion of alphabets takes place e.g. 'LISP-LIPS'. They occur mainly due to learning disabilities or neurological disorders (mental retardation, cerebral palsy, neonatal jaundice) [1].

Methodology for building the system can be described below

8.1 Picture naming task

The person suffering from articulation errors speak a particular word into the built system. For this purpose a picture naming task can be made. They can be asked to look at a picture and recognize the word or the word itself can be written on the picture and they can read it aloud into the system.

8.2 Recording the response

The recordings must be taken using efficient software (wavepad, goldwave or praat). The recordings should be in silent environment for optimal results [1].

8.3 Preprocessing

The recorded sound often consists of unwanted sounds or disturbance regions. These must be removed to improve the accuracy of the system.

8.3.1 Speech enhancement

Sometimes, the sound is degraded by noise, so it must be enhanced. Different speech enhancement techniques have been compared i.e. Weiner filtering, spectral subtraction, log MMSE and it has been concluded that one of the best speech enhancement tools analyzed is log minimum mean square error (log MMSE). Weiner filtering works well only when there are no musical distortions. MMSE technique requires information about signal and noise spectrums beforehand. Signal to noise ratio is calculated by SNR estimator. The spectral is analyzed by calculating mean square error in between clean and estimated spectral [1].

8.3.2 Silence removal

It refers to remove silence regions from speech (enhanced one). For this purpose short term energy of signal and centroid (center of the spectral) are calculated. Energy and centroid of voiced speech is greater than that of unvoiced or silence region in a speech. So, they are filtered using median filter method. These two features are firstly extracted and there threshold is determined [1] [6].

8.4 Feature extraction

Extraction is done to extract feature vectors (only the useful features) from the speech signal obtained. This can be done by following techniques

8.4.1 Framing

Speech signals are quasi-stationary (almost stationary) by their behaviour. So, it's very difficult to use the whole signal at once. It's better to divide the signal into segments (frames) and then process them individually. For this purpose framing is done. Length of a typical frame is 20-30 ms. these may overlap by 40-60% so as to prevent the loss of signals when we apply windowing over it. Number of samples (in a frame) = Sampling rate (of the frame) * Duration (of each frame).

8.4.2 Windowing

It is done by applying Hamming / Hanning window functions. The main purpose of windowing is to minimize the effect of signal loss (attenuation) produced during framing at beginning and end point of every single frame.

Hanning window can be calculated by

$$W(x) = 0.5\,(1 - \cos(2\pi x / Y)),\ 0 <= x <= Y \qquad (2)$$

where, x is a block in frame and Y is the total number of samples present in frame. Apart from these various feature extraction techniques can be used. (Cepstral Analysis, Mel frequency cepstral coefficients, Linear Prediction Coding, Perceptually based Linear Predictive analysis, Principle Component Analysis). Some of them have been explained further.

8.4.3 MFCC (Mel frequency Cepstral coefficients)

Figure2. defines the process used in MFCC [3].

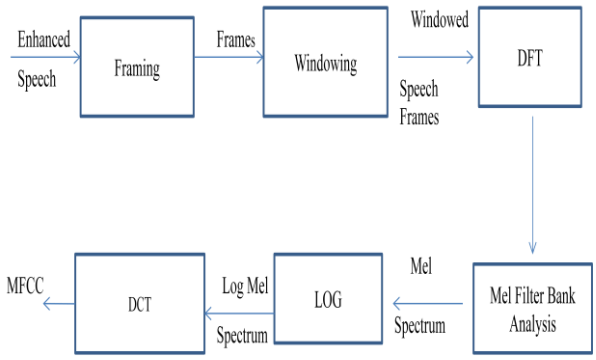

Fig 2: MFCC Extraction [3]

First two steps are common. Framing and Windowing can be done and feature vectors can be extracted. After that, short term analysis is done and MFCC feature vector is calculated.

Steps for MFCC extraction

- *Calculate Periodogram based power spectral analysis*

Apply discrete Fourier transformation series on each signal frame and spectral estimate of each frame is calculated.

- *Mel frequency scale warping*

Periodogram contains lots of unnecessary information which is needed to be warped off (removed). This is done by Mel Frequency Scale thus obtaining the filter bank.

MFCC feature vector can be calculated as

$$Mel\,(f) = 2595 * \log10\,(1 + f/700) \qquad (3)$$

where, Mel(f) is the mel filter bank. Mel filter bank contains collection of overlapping triangles called Mel Spaced.

- *Log Mel Spectrum*

Applying log on Mel spectrum so that output should be in compressed form to watch it more closely in relation with human auditory system.

- *Applying Discrete cosine transformation series*

DCT converts log Mel spectrum to spatial domain. The results thus obtained after applying cosines are MFCC (Mel frequency Cepstral coefficients).

- *Delta energy and Delta spectrum*

It increases accuracy of MFCC. Delta is first order cepstral derivative while delta-delta is second order cepstral derivative. After all this, normalization of vectors takes place to get most accurate results.

8.4.4 Linear prediction Cepstral coefficient (LPCC)

LPC is used in the first step to obtain LPCC coefficients. Then they are converted into Cepstral coefficients. It uses autocorrelation method [10].

8.4.5 Linear Prediction Coding

Linear prediction coding is a powerful method for linear prediction. It provides estimation of speech in a précised manner. Its main aim is to analyze past speech samples. Steps involved in LPC are shown in figure3 [3].

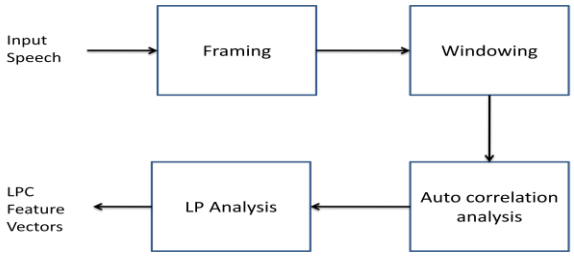

Fig 3: LPC Extraction

8.4.6 Perceptually Based Linear Predictive Analysis (PLP)

It is mainly for cross-speaker isolated word recognition. It has two steps obtaining auditory spectrum (derived by critical band filtering of speech waveform) and applying spectrum to the all pole model. It provides same results as LPC except that its order is half then that of LPC. Thus, saving storage in ASR and reducing its computation. The process of PLP is shown in figure4 [10].

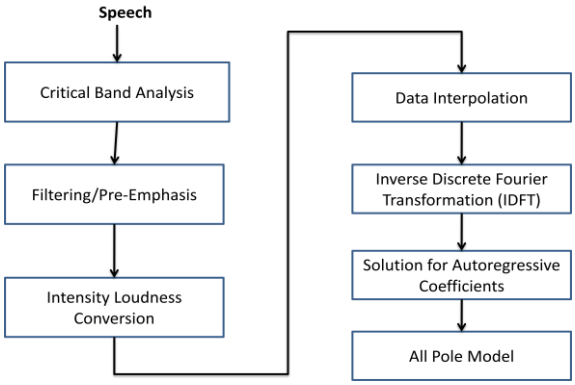

Fig 4: PLP Extraction

These are few feature extraction techniques. A comparison between all these has been done in Table 4 [22][23][24][25][27].
Any of these can be implemented for feature extraction to train articulation error detection model [10].

Table 4. Feature Extraction Methods

Method	Property	Description
Linear Predictive Coding	Static Method	Feature Extraction at lower order
Cepstral Analysis	Static, Spectral Analysis	Spectral Analysis is performed by Mel-frequency scale
Principal Component Analysis (PCA)	Non-Linear, fast	Based on Eigenvectors, Good for Gaussian data
MFCC	Discrete Fourier Transform Series	Framing, Windowing, Cosine Transformations
Linear discriminant Analysis (LDA)	Non-Linear, Supervised, fast	Based on eigenvectors, Better than PCA
Fusion MFCC	MFCC+LDA	Better results in continuous speech recognition system
Kernel-Based	Non-Linear	Removes redundant and noisy data, Reduction in dimensions leads to better classification
Wavelet Method	Better then DFT	Better time resolution then Fourier
Spectral Subtraction	Robust	Forms and work on Spectrograms
RASTA	For Noisy speech	Finds Noisy (data) features
Integrated method	PCA+LDA+ICA	High accuracy
Cepstral mean subtraction	Robust	Similar to MFCC, works on mean static parameters
PLP(Perceptually Based Linear Predictive Analysis)	Similar to LPC except that it's order is half then that of LPC	Saving storage in ASR and reducing computation

9. MODULES OF ASR
The main modules of automatic speech recognition is described below

9.1 Speech signal acquisition
Speech is recorded by the picture naming task (explained in section 8). It is then converted to digital form for further analysis.

9.2 Feature Extraction
Various feature extraction schemes, explained before can be used as per requirement and accuracy. MFCC is one of the best techniques till date.

9.3 Acoustic Models
Acoustic modeling is done to develop link between speech signal and expected output (word, sentence or in this case an articulation error). The system is thus rigorously trained so as to get the desirable output.

9.4 Language Models and Lexical Models
The issues like word ambiguity and boundaries are handled by language models. Language model generates probability of all the possible words. Lexical models on the other hand handles the pronunciation factor e.g. how the word should be spoken. Each word has its own way of speaking. How they should be spoken in a system is decided by lexical models.

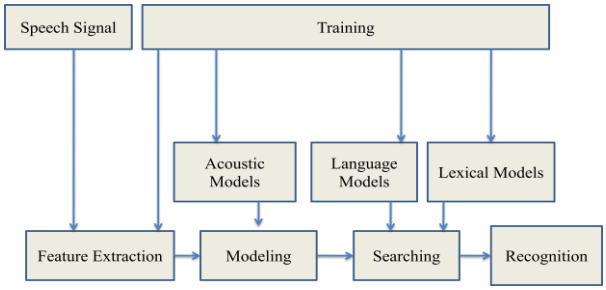

Fig 5: Modules of ASR

This is the main process in building a speech recognition system. (Figure 5) [4][7].

9.5 Model Adaptation
It decides which model to apply for what kind of domain. The models that can be implemented to build the system have been discussed in next section. It can be chosen according to specification and requirements.

9.6 Recognition
Recognition is the main phase in ASR. It decides whether the built system is providing desired results or not. The system built for articulation error detection should be able to recognize the error. The unknown pattern is compared to reference pattern in case of pattern matching model which is discussed in the next section [2] [4].

10. SPEECH RECOGNITION TECHNIQUES
Various speech recognition techniques have been explained below. Amongst these, the approach used for articulation error detection will be pattern recognition approach (figure 6).

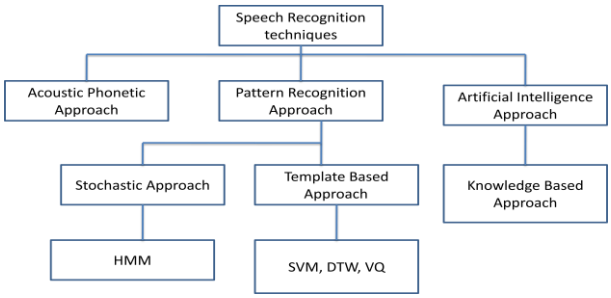

Fig 6: Speech recognition techniques

Pattern approach can be further classified into model based approach, template approach and classified approach such as HMM, DTW, VQ, SVM. A brief description about all these techniques is explained further [2] [5].

10.1 Acoustic Phonetic Approach
There exists distinctive and finite phonemes in speech and they are characterized by their acoustic properties. The

foremost step in building this model is spectral analysis of speech. The next step is conversion of spectral into features which defines acoustic properties of various phonemes. Then segmentation and labeling is performed where the signal is segmented into isolated regions. The last stage is finding the articulation error from labels. This technique has not been used much [2] [5] [34].

10.2 Pattern Recognition Approach

It involves two steps Pattern training and Pattern matching (to check whether they match or not). It is one of the most used techniques and its reliability factor is also high. It uses mathematical framework to represent speech patterns which increases its reliability and accuracy. It even matches the unknown speech (not present in the training set) and predicts the desired output. It involves a four stage process mainly feature extraction, training of pattern, classification of pattern and decision logic. Thus, unknown pattern and reference pattern are compared and distance (between them) is computed. At last, Decision logic decides how much they (the patterns) match [2]. These are of two types template based and stochastic approach.

10.2.1 Template Based Approach

A template is made of known speech pattern and an unknown speech pattern is matched with it so as to identify the best match possible. It is one of the most used approaches today. The dictionary made with known speech is referred as reference pattern. Unknown speech is matched to it and threshold of similarity is obtained. Major disadvantage in this kind of approach is that the variation in speech is calculated by many templates a word which becomes unfeasible after a while [2] [5] [29].

10.2.2 Stochastic Based Approach

Stochastic approach gives result even if the information provided to it is incomplete. One of the major examples of this approach is HMM (HIDDEN MARKOV MODEL) which is explained further [30] [31].

10.2.2.1 Hidden Markov Model (HMM)

Hidden Markov model abbreviated as HMM contains finite set of State Markov model as input and a set of output. Parameters in the input are referred as Temporal Variability while those in output are called Spectral Variability. The main problem with HMM is the determination of best state of model (solved by understanding physical meaning of the states of model) and their adjustments for the best account of observed signals. Also, finding probability for the observations is a tough task. HMM involves the following steps

- Define no. of sounds e.g. S, these can be words also
- Define classes as K = $\{k_1, k_2 \ldots . k_S\}$
- Form a training set with collected sounds
- Solve estimation problem for class K
- Evaluate P (O/λ) (i = 1, 2, . . . , S) where, λ is the best model for each class
- Identification of speech O of class K

10.2.3 Dynamic time warping (DTW)

DTW is used to measure similarity in two sequences (speech) which varies in speed and time, by finding the optimal match between them. (e.g. if in a video a person is walking slow and in the other he is walking fast, the patterns can still be matched by DTW). Continuity in signals is not much of an importance in case of DTW. It is applied for linear sequences mainly (audio, video or graphics) [28]. It matches sequences with missing information. Optimization process is particularly done by dynamic programming and this is why it is named DTW. Very less work has been done in DTW [3] [11]. DTW has two types of which, one is symmetrical in which every frame in the input pattern is used in the matching path and another is asymmetrical DTW in which the frames are used only once from the input pattern and cannot be repeated again [32].

10.2.4 Vector Quantization (VQ)

Vector Quantization is used for data reduction. It mainly uses compact codebooks as reference models and codebook searchers rather than high cost evaluation methods. The word is chosen by the system which gives minimum distance. The entries in the codebook are not ordered. (Order not maintained in training and reference pattern data). It divides a set of vectors or points into groups having same number of vectors close to them. Each group is represented by its center. VQ can also be used in case of density estimation and lost data correction. The minimum distance between a vector and the codebook is known as VQ Distortion. In recognition phase, total distortion is calculated. Data from the speaker is matched in the codebook and difference is measured which in turn is used to make the decision of recognition or in this case detection of errors [3][11][12].

LBG: Linde-Buzo-Gray (LBG):

LBG is designed in the community of VQ for data compression. A speaker can be differentiated from others on the basis of location of centroid. ‘m’ are the training vectors and ‘M’ are codebook vectors in Figure7 [13] [11].

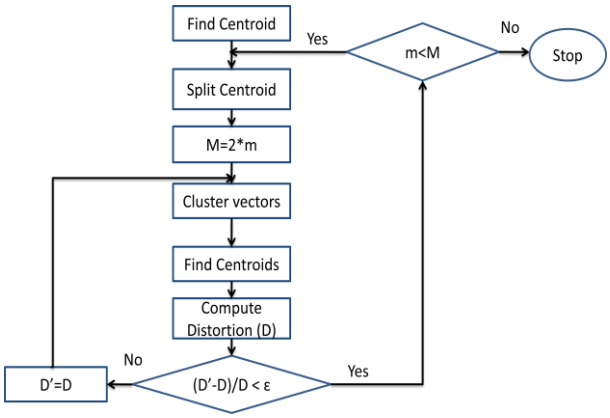

Fig 7: LBG Flowchart

10.3 Artificial Intelligence Approach (Knowledge based approach)

Artificial Intelligence refers to intelligence of a person (analyzing and visualizing) used in decision making. It is a mixed form of Acoustic and Pattern approaches. Rules and procedures are applied and knowledge makes algorithms much better. The problem of Artificial Intelligence lies in, it is not wholly successful due to the complications raised while quantifying expert knowledge. The second problem lies in the fact that it integrates levels of human knowledge such as syntax, lexical analysis and semantics [14] [27].

10.4 Artificial Neural Networks

ANN is also known as Connectionist Approach. In this model knowledge is not represented by rules or procedures, rather it is distributed in computing units which is simple by nature. Because the computation has some similarity to nervous system, these models are known as artificial neural networks. Training can be expensive and data may require large number of iterations. Still the simplicity and uniformity of underlying processes can make this model attractive to implement [15].

10.5 Support Vector Machine (SVM)

It is an approach of pattern recognition that uses discriminative approach. SVM uses hyper-planes for classification of data. It is independent of dimensions and uses large dimension space to construct huge number of non-linear feature vectors. After this, during the training phase, feature selection is performed. SVM uses boundaries to differentiate between classes and objects. It maps input using kernel functions and classes are separated using hyper-plane classification. It is also used for emotion recognition and voice activity detection apart from speech recognition. By using radial basis functions and LIBSVM implementation can make it more efficient. Articulation error detection has been successfully done with this and yields an accuracy of 92%. Table 5 shows various speech recognition techniques along with their recognition functions [36],

Table 5. Speech Recognition Approaches

Techniques	Representation	Recognition Features
Acoustic Approach	Spectral Analysis with error detection phonemes	Probabilistic analysis
Pattern recognition approach Template, DTW, VQ	Set of spectral vectors and Features	Minimal distance measured Dynamic Warping Optimal Clustering algorithm
Neural Network	Perceptrons/Units	Network/Activation Functions
SVM	Kernel Based/Radial Basis Functions	Hyper-Plane Classification
AI approach	Knowledge Base	Rules/Procedures/Algorithms/ Word error probability

11. COMPARATIVE ANALYSIS

A brief description about modern day changes done so far in the field of speech recognition is represented in Table7. Technology is at its peak now. Large vocabulary based, independent system has been built. To add more features to it, it can be related to articulation error detection and correction tool [4] [5] [18] [33].

Table 6. Comparative Analyses

Earlier	Today
Distance-Based Models	Likelihood Based Models
Isolated systems	Connected word recognition
Small Vocabulary	Large Vocabulary
Single Speaker	Adaptive Recognition
Normalization of feature vectors	DTW Approach
Noiseless Speech Recognition	Noisy Speech Recognition
Template matching	Stochastic approach (HMM)

12. TOOLS FOR SPEECH RECORDING

Many tools are available as open source for recording purpose. For creating the database, most important thing is to download efficient software for recording purpose. The software should be user friendly and should provide sharp noise signals [4]. Goldwave is best amongst all these as it support features like recording from multiple sources, editing and giving special effects. Table7 contains a brief description of all these tools [19] [20] [21].

Table 7. Tools Used

NAME OF THE TOOL	SOURCE	LATEST VERSION	DESCRIPTION
AUDACITY	OPEN	1.3.14 (Beta)	Editing and Recording Sounds
CSL	OPEN/ PROPRIETARY	Model 4500/4150 B	Data Acquisition, Speech Analysis
GOLDWAVE	OPEN	Version 6.0	Recording and Speech Analysis
HTK	OPEN	3.4.1	Written in ANSI C, Build and manipulate HMM
PRAAT	OPEN	5.3.04	Recording and Speech Analysis in stereo or mono
SCARF	OPEN	2009	Speech Recognition with Segmental conditional random fields
SPHINX	OPEN	2.2.10, SPHINX4	Written in JAVA, Analysis of speech
WAVEPAD SOUND EDITOR	OPEN	6.34	Recording and Speech Analysis

13. CONCLUSION

There are several ways to help children/adults suffering from speech disorders. Few of them are to create an accurate speech model, so that the persons can check whether they have disorders or not and be a knowledgeable referral source to understand the developmental norms. Articulation therapies can be provided which starts from making children learn the syllables first, words next, then sentences and further making them learn stories and generalization. Speech models can be built using HMM, VQ techniques along with MFCC and advanced MFCC as feature extraction for accuracy. Further, after building this an error treatment tool can also be made for different languages. This can help a person suffering from articulation error correcting his/her sound.

14. REFERENCES

[1] Singh, Shailendra, Anshul Thakur, and Dharam Vir. "Automatic articulation error detection tool for Punjabi language with aid for hearing impaired people." *International Journal of Speech Technology* 18.2 (2015): 143-156.

[2] Bhabad, Sanjivani S., and Gajanan K. Kharate. "An Overview of Technical Progress in Speech Recognition." *International Journal of Advanced Research in Computer Science and Software Engineering* 3.3 (2013).

[3] Radha, V., and C. Vimala. "A review on speech recognition challenges and approaches." *doaj. Org* 2.1 (2012): 1-7.

[4] Arora, Shipra J., and Rishi Pal Singh. "Automatic speech recognition: a review." *International Journal of Computer Applications* 60.9 (2012): 34-44.

[5] Anusuya, M. A., and Shriniwas K. Katti. "Speech recognition by machine, a review." *arXiv preprint arXiv: 1001.2267* (2010).

[6] Natarajan, V. Anantha, and S. Jothilakshmi. "Segmentation of continuous speech into consonant and vowel units using formant frequencies."*International Journal of Computer Applications* 56.15 (2012).

[7] Ghai, Wiqas, and Navdeep Singh. "Analysis of automatic speech recognition systems for indo-aryan languages: Punjabi a case study." *Int Journal of Soft Computing* 2.1 (2012): 379-385.

[8] Muda, Lindasalwa, Mumtaj Begam, and I. Elamvazuthi. "Voice recognition algorithms using mel frequency cepstral Coefficient (MFCC) and dynamic time warping (DTW) techniques." *arXiv preprint arXiv: 1003.4083* (2010).

[9] Das, Sanjib. "Speech recognition technique: A review." *Int Eng Res Appl*2.3 (2012): 2071-2087.

[10] Shanthi Therese, S., and Chelpa Lingam. "Review of Feature Extraction Techniques in Automatic Speech Recognition." *International Journal of Scientific Engineering and Technology* 2.6 (2013): 479-484.

[11] Saini, Preeti, and Parneet Kaur. "Automatic speech recognition: A review."*International journal of Engineering Trends & Technology* (2013): 132-136.

[12] Gaikwad, Santosh K., Bharti W. Gawali, and Pravin Yannawar. "A review on speech recognition technique." *International Journal of Computer Applications* 10.3 (2010): 16-24.

[13] Shanthi Therese, S., and Chelpa Lingam. "Review of Feature Extraction Techniques in Automatic Speech Recognition." *International Journal of Scientific Engineering and Technology* 2.6 (2013): 479-484.

[14] Gaikwad, Santosh K., Bharti W. Gawali, and Pravin Yannawar. "A review on speech recognition technique "*International Journal of Computer Applications* 10.3 (2010) : 16-24

[15] Desai, Nidhi, Kinnal Dhameliya, and Vijayendra Desai. "Feature extraction and classification techniques for speech recognition: A review."*International Journal of Emerging Technology and Advanced Engineering* 13.12 (2013): 367-371.

[16] Srinivasan, A. "Speech recognition using Hidden Markov model." *Applied Mathematical Sciences* 5.79 (2011): 3943-3948.

[17] Yu, Youhao. "Research on speech recognition technology and its application." *2012 International Conference on Computer Science and Electronics Engineering*. IEEE, 2012.

[18] Yu, Youhao. "Research on speech recognition technology and its application." *2012 International Conference on Computer Science and Electronics Engineering*. IEEE, 2012.

[19] Sharma, Vivek, and Meenakshi Sharma. "A quantitative study of the Automatic Speech Recognition Technique." *International Journal of Advances in Science and Technology* 1.1 (2013).

[20] Luthra, Simmi, and Parminder Singh. "Punjabi Speech Generation System based on Phonemes." *International Journal of Computer Applications* 49.13 (2012): 40-44.

[21] Aggarwal, Naveen. "Analysis of Various Features using Different Temporal Derivatives from Speech Signals." *International Journal of Computer Applications* 118.8 (2015).

[22] Al-Barhamtoshy, Hassanin, et al. "Speak Correct: Phonetic Editor Approach." *Life Science Journal* 11.8 (2014).

[23] Strik, Helmer. "ASR-based systems for language learning and therapy." (2012).

[24] Lee, Sungjin, et al. "Grammatical Error Detection for Corrective Feedback Provision in Oral Conversations." *AAAI*. 2011.

[25] Rughani, Megha, and D. Shivakrishna. "A Review on Dysarthric Speech Recognition." *International Journal of Advanced Networking & Applications* (2014).

[26] Dixit, Ranu, and Navdeep Kaur. "Speech Recognition Using Stochastic Approach: A Review." *International Journal of Innovative Research in Science, Engineering and Technology* 2.2 (2013).

[27] Bertucci, Carol, et al. "Vowel perception and production in adolescents with reading disabilities." *Annals of Dyslexia* 53.1 (2003): 174-200.

[28] Aihara, Ryo, et al. "A preliminary demonstration of exemplar-based voice conversion for articulation disorders using an individuality-preserving dictionary." *EURASIP Journal on Audio, Speech, and Music Processing*2014.1 (2014): 1-10.