International Journal of Advanced Technology in Engineering and Science
Vol. No. 10, Issue No. 06, June 2022
www.ijates.com
ijates
ISSN 2348 - 7550

A REVIEW ON ARTICULATORY KNOWLEDGE IN THE RECOGNITION OF DYSARTHRIC SPEECH

Md. Nadeem Enam [1], Ozair Ahmad [2], Dr. Tajuddin Ali Ahmad[3], Md. Naushad Akhtar[4]

[1,2,3,4] *Assistant Professor, Department of ECE, Maulana Azad College of Engineering & Technology, Patna, (India)*

ABSTRACT

Dysarthria is a motor speech disorder in which the muscles that are used to produce speech are damaged, paralyzed or weakened. Dysarthria often causes slurred or slow speech that can be difficult to understand. Dysarthric speech led to impairments in intelligibility, audibility, naturalness and efficiency of vocal communication.

Articulation is the process by which we produce sound and it involves lungs, vocal tracts, jaws, facial muscles, lip and tongue. The articulators (set of organs) are not correctly positioned in case of dysarthric speech which results in the production of unintelligible sound. Hence, articulatory knowledge in the recognition of dysarthric speech play a very pivotal role.

In this paper there is a complete review on articulatory knowledge in the recognition of Dysarthric speech.

Keywords: Dysarthria, Speech Recognition, Articulatory Models, Bayesian Network, Articulatory Information

I. INTRODUCTION

Millions of people are suffering from congenital speech disorder in which they don't have complete control over their muscles and as a result unable to produce intelligible/ controlled sound. These conditions are called dysarthria and production of unintelligible/un controlled speech are known as Dysarthric speech. Dysarthria can be characterized in terms of poor control over articulatory movement. Anything that causes brain damage can cause dysarthria viz. stroke, brain injury, tumors, Cerebral palsy etc.

Dysarthria affects the articulation. It can be mild or severe. So, in this circumstance, these speeches are very difficult to understand not only by the human being but also by the automatic speech recognition (ASR). Articulatory features are quite helpful in these conditions which help out in parameterized the articulation phenomena [1].

In the recent past, Speech recognition technology have considerably changed. The research on human speech

production in Automatic speech recognition has taken different dimensions right from the simple combination of articulatory and acoustic features to complex hidden dynamic models of articulatory movements [2].

Several attempts were made to improve speech recognition as well as integration of articulatory knowledge for a dysarthric person. We require an automatic speech recognition frame work based model for studying the dynamics of dysarthric speech [1].

II. LITERATURE REVIEW

This section elaborates the previous work performed by the researcher on articulatory knowledge in the recognition of dysarthric speech.

1.1 Recognition of dysarthric speech

In all the previous development in Automatic speech recognition work for individuals with dysarthria was almost exclusively centered towards Hidden Markov model (HMM). This model was extremely important as because its parameter was trained for the general population. The Word recognition rates were lie somewhere in between 26 % to 82% which is more often lower for a dysarthric speakers than a healthy speaker. Researchers had also found that the percentage of voice recognition by HMM model for a severely dysarthric speaker (affected by cerebral palsy) was merely around 4%. Whereas this percentage was substantially high (nearly 89%) for non-dysarthric speakers. So, they had found a huge disparity in the process of modeling of data through HMM model. Although the dysarthric speeches do not exhibit due to a single reason, rather than it may be a combination of several articulatory features [1],[16],[17].

Polur and Miller found in his research work that ergodic HMM allowed backward transitions which eventually resulted in the slight improvement of the dysarthric speech results. Further, Raghavendra compared a speaker adaptive phoneme recognizer with a speaker dependent word recognizer for the different dysarthric speech and found that the voice recognition adaptation is relatively better in case of mild to moderate level of dysarthric speeches which was having a relative error reduction (RER) of 22% to 47% near about. It was a significant work by him which was later on more improved in terms of improving the relative error reduction by using weighted transducer in automatic speech recognition systems [4].

Accuracy for dysarthric speakers have been increased to some extent with the help of adaptation at the acoustic level. Since dysarthria was closely associated with the speech production mechanism, in spite of this fact the researchers have not focused on the physiological model [5],[18].

1.2 Articulatory knowledge in speech recognition

Several attempts were made by researcher to develop theoretical production knowledge directly into models for speech recognition. Empirical articulatory measurement to acoustic observations has been shown to reduce phone error in a standard Hidden Markov model, further it has also been noticed that articulatory measurement was disimproved due to the interference of acoustic.

Hence, researchers were continuously trying to conduct the experiment for the system which could produce an improved result in the presence of acoustic. Eventually they became successful to correlate the system that can

learn discrete articulatory feature from neural network in the presence of acoustics and able to fit or model these data into a standard HMM model which showed some improved results. However, the importance of this result was not always statistically significant as there was always a possibility of environmental noise to interfere with the system [1] [3]

In a similar baseline Metze found that with the inclusion of discrete articulatory features gained from maximum mutual information could reduce the word error rates by a fair amount [1].

Hence, we can say that the use of direct articulatory knowledge can sufficiently reduce the word errors in the recognition of speech. It becomes more effective if the articulatory knowledge is influenced by high level abstraction of vocal tract behavior.

In recent days, Bayes networks have been used in modelling interdependencies between articulation and acoustics in a regular speech. Bayes networks that estimated the likelihood of acoustic observations given discretized articulatoryparameters, achieving similar results when combined with an HMM- based ASR system [4].

1.3 Articulatory Features/ Diagram

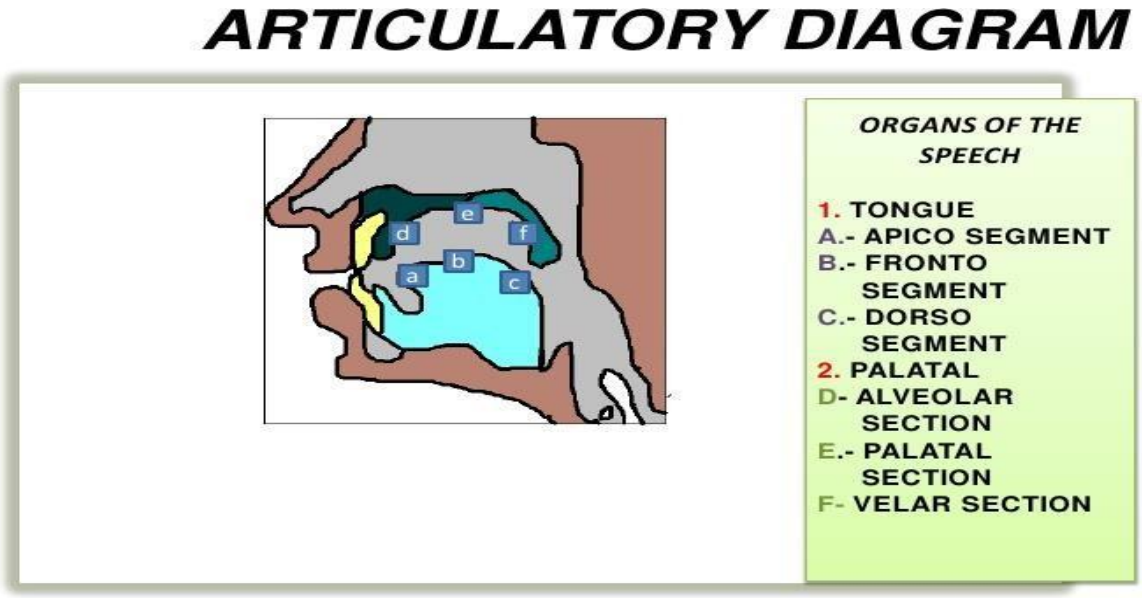
Features are nothing but logic or numbers. In other word we can say that features are those parameters with the help of which we can define any physical phenomenon. Features should be self- sufficient.

Hence, articulatory features can be defined as numbers which help in parameterized the articulation phenomena. Dysarthria affects the articulation. So articulatory features play an essential role in order to recognize the dysarthric speech from ASR systems. Articulatory Features might be useful for adaptation, particularly to speaker and speaking style as because they have been used in their respective analysis [9], [13].

The main articulators are the tongue, the upper lip, the lower lip, the upper teeth, the upper gum ridge, the hard palate, the soft palate and the glottis (space between the vocal cords). Tongue is the most important articulator of speech. The articulators are used to change the properties of the acoustic filter over the time. Articulatory phonetics is the study of the way the vocal organs are used to produce speech sound, whereas the acoustic phonetics is the study of the physical properties of the speech sounds. In fact, phonetics and phonology differs from each other [15], [19].

Phonetics deals with how speech sounds are actually produced whereas phonology deals specifically with the ways those sounds are organized into the individual language [10][14].

The articulatory diagram is shown in the figure given below.



1.4 Representations for speech production knowledge

A more sophisticated approach to vocal tract knowledge has been derived from actual measurement of the vocal tract during speech with semi- invasive procedures viz. electromagnetic articulography (EMA), magnetic resonance imaging (MRI) X-ray microbeam analysis, ultrasound etc.

Little amount of current is induced into small receiver coils glued to the jaws, tongue, lips, fascial muscles and other articulators. In this process, the positions of these articulators can be accurately deduced relative to the fixed transmitters around the speaker's head that produce alternating magnetic fields. Production of audible noise is almost nil in these systems and hence the coils interfere surprisingly little with continuous speech [2]. Figure 1 shows the typical configuration of EMA cube with the respective placement of the receiver coils [6],[8].

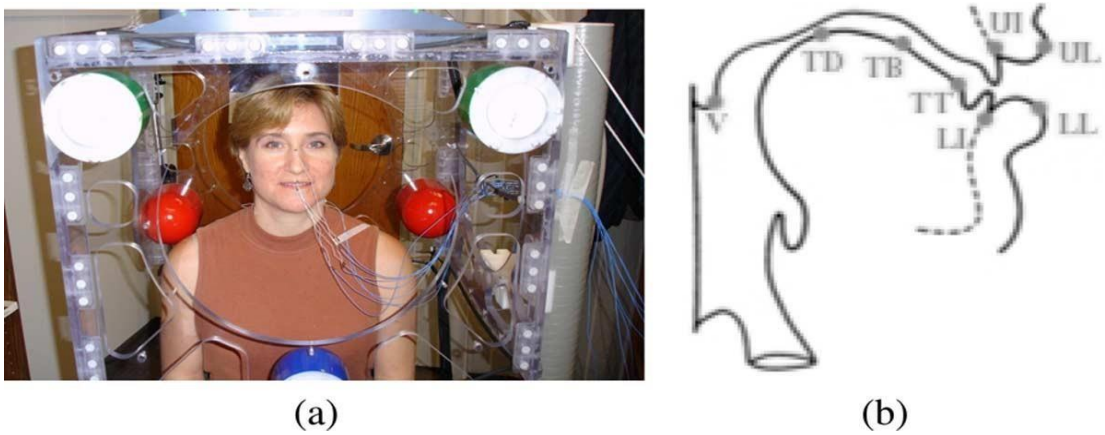


Fig.1.Example configuration of electromagnetic articulography. (a) shows a subject connected within the recording environment, and (b) shows the typical locations of receiver coils.

Articulatory features have been collected into seven different categories. Every single category has been assigned by a unique number. For instance, a segment of speech may be concurrently voiced, nasal and static which represents the values in number for three different features. Some other very useful attributes of articulatory features (AFs) comprise of language independencies and trust worthy recovery from acoustics among regular speaker. In the absence of articulatory feature annotations, AF values can be obtained directly from the phoneme annotations. Many a times articulatory features are also known as phonological features.
Some articulatory features and their respective descriptions have been pointed out in the following table I [4].

ARTICULATORY FEATURES, DESCRIPTION OF THEIR CHARACTERISTICS AND THEIR POSSIBLE VALUES.

Feature	Description (*and values*)
Manner (**M**)	high-level categorization of speech sound *approximant, fricative, nasal, retroflex, silence, stop, vowel*
Place (**Pl**)	location of primary constriction *alveolar, bilabial, dental, labiodental, velar, silence,nil*
High/Low (**HL**)	ventral position of the tongue *high, mid, low, silence, nil*
Front/Back (**FB**)	anterior position of the tongue *front, central, back, nil*
Voice (**V**)	presence/absence of glottal vibration *voiced, unvoiced*
Round (**R**)	circularity of the lips *round, non-round, nil*
Static (**S**)	movement of articulators (e.g., diphthong) *static, dynamic*

Table 1

In the above table each frame of data has been assigned a seven-dimensional vector of AF values based exclusively on phoneme annotation at that frame.

1.5 Dynamic Bayesian network (DBN)

During speech recognition, it was very tough job to get the observations of articulator's movement. Thus, researcher came out with the innovative idea to focus largely on modeling the hidden articulatory trajectory with prior phonetic knowledge and hence found that the combination of articulatory information with acoustic lead towards a better speech recognition performance.
Bayesian networks (BN) which is based on probabilistic dependency approach in between articulatory features and its corresponding acoustic features are highly suitable for speech recognition. BN networks have the flexibility to model the complex joint PDF with many continuous and discrete variables. Speech recognition systems are basically hidden Markov models (HMM/BN) where BN is used to define the HMMs probability distributions. In the process of recognition, articulatory variables are assumed hidden during the presence of acoustic observations only [25]. A small data base consisting of articulatory and acoustic observation recorded

from three speakers have been used for the performance evaluation of the Automatic speech recognition system [2].

With the help of dynamic Bayesian network, we can extract the articulatory features. We can increase the recognition accuracy by modeling the dependencies between a set of 6 multi-leveled articulatory features, over an equivalent system in which all the features are assumed to be independent. Bayesian network (BN) actually provides a tool for the encoding of dependencies between a set of random variables (RV). We can represent these random variables and dependencies in the form of a directed graph where random variables are designated as node and the dependencies through edges [3],[6],[15].

In the Figure 2, typical Bayesian network model has been shown.

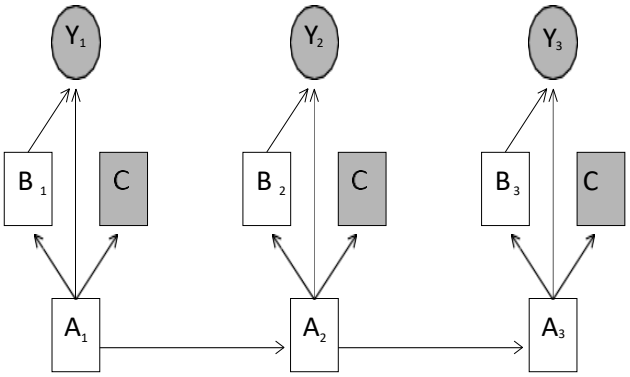


Figure 2

Mathematically, we can represent this dynamic Bayesian network (DBN) in terms of joint probabilities as given below.

$p(A1, B1, C1, Y1)$

$= p(Y1|A1, B1)\ P(B1|A1)\ P(C1|A1)\ P(A1)$

where $P()$ and $p()$ denotes probability mass function and probability density function respectively.

Dynamic Bayesian network forms a large class of models which provide an ideal framework to integrate the information from various available sources. The motive behind this work is a word recognizer built around an articulatory feature separating the state and observation processes. [3],[20],[21].

REFERENCES

[1] Frank Rudzicz " Using articulatory likelihoods in the recognition of dysarthric speech "Speech communication 54 (2012) 430-444

[2] Konstantin Markov, Jianwu Dang and Santoshi Nakamura, " Integration of articulatory and spectrum features based on the hybrid HMM/BN modeling framework "Speech communication, vol.48, no. 2, pp. 161-175, February 2006

[3] Frankel, Miriam wester, and Simon King," Articulatory features recognition using dynamic Bayesian networks", Computer speech and language, vol.21, pp 620-640,2007

[4] Frank Rudzicz " Articulatory knowledge in the recognition of dysarthric speech" IEEE Transactions on Audio, speech and language processing, Vol-19, no.4, May,2011

[5] Frank Rudzicz " Applying discretized articulatory knowledge to dysarthric speech", 978- 1-4244-2354-5/09,2009, IEEE

[6] Kevin Petrik Murphy, Dynamic Bayesian Networks: Representation, Interference and Learning, PhD thesis, University of California at Berkeley,2002

[7] Karen Livescu, Ozgur Cetin, Mark Hasegawa- Johnson, Simon King, Chris Bartels, Nash Borges, Arthur Kantor, Paratha Lal, Lwisa Yung, Ari Bezman, Stephen Dawson- Haggerty and Born- woods " "Articulatory feature-based methods for acoustic and audio- visual speech recognition: Summary from the 2006 JHU Summer Workshop," in Proceedings of ICASSP 2007, Honolulu, April 2007

[8] John-Paul Hosom, Alexander B. Kain, Taniya Mishra, Jan P. H. van Santen, Melanie Fried- Oken, and Janice Staehely, "Intelligibility of modifications to dysarthric speech," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '03), April 2003, vol. 1, pp. 924–927.

[9] Miriam wester, "Syllable classification using articulatory-acoustic features", in proceedings of Euro speech 2003, Geneva, Swizerland,2003, pp-233-236.

[10] M. Hasegawa-Johnson, J. Gunderson, A. Perlman, and T. S. Huang, "Audio visual phonologic-feature-based recognition of dysarthric speech," Abstract, 2006.

[11] K. Rosen and S. Yampolsky, "Automatic speech recognition and a review of its functioning with dysarthric speech," Augment. Altern. Common. vol. 16, no. 1, pp. 48–60, Jan. 2000 [Online]. Available: http://dx.doi.org/10.1080/07434610012331278904

[12] K. Kirchhoff, "Robust speech recognition using articulatory informa- tion," Ph.D. dissertation, Univ. of Bielefeld, Bielefeld, Germany, July 1999

[13] F. Metze, "Discriminative speaker adaptation using articulatory features," Speech Comm., vol. 49, no. 5, pp. 348–360, 2007.

[14] N. Borges, A. Kantor, P. Lal, L. Yung, A. Bezman, S. Dawson- Haggerty, and B. Woods, "Articulatory feature-based methods for acoustic and audio-visual speech recognition: Summary from the 2006 JHU summer workshop," in Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2007), Honolulu, HI, Apr. 2007, pp. IV-621–IV-624

[15] M. Wester, J. Frankel, and S. King, "Asynchronous articulatory fea- ture recognition using dynamic Bayesian networks," in Proc. Inst. Electron., Inf. Commun. Eng. Beyond HMM Workshop, Kyoto, Japan, 2004, vol. 104, pp. 37–42.

[16] C. Havstam, M. Buchholz, and L. Hartelius, "Speech recognition and dysarthria: A single subject study of two individuals with profound im- pairment of speech and motor control," Logopedics Phoniatrics Vo- cology, vol. 28, no. 10, pp. 81–90, Aug. 2003.

[17] E. Sanders, M. Ruiter, L. Beijer, and H. Strik, "Automatic recognition of Dutch dysarthric speech: A pilot study," in Proc. 7th Int. Conf. Spoken Lang. Process., Denver, CO, Sep. 2002.

[18] S. King, J. Frankel, K. Livescu, E. McDermott, K. Richmond, and M. Wester, "Speech production knowledge in automatic speech recognition," J. Acoust. Soc. Amer., vol. 121, no. 2, pp. 723–742, Feb. 2007

[19] Wrench and K. Richmond, "Continuous speech recognition using articulatory data," in Proc. Int. Conf. Spoken Lang. Process., Beijing, China, 2000

[20] K. Markov, J. Dang, and S. Nakamura, "Integration of articulatory and spectrum features based on the hybrid HMM/BN modeling frame- work," Speech Commun., vol. 48, no. 2, pp. 161–175, Feb. 2006.

[21] K. P. Murphy, "Dynamic Bayesian networks: Representation, infer- ence and learning," Ph.D. dissertation, Univ. of California at Berkeley, Berkeley, CA, 2002.

[22] J.-P. Hosom, A. B. Kain, T. Mishra, J. P. H. van Santen, M. Fried-Oken, and J. Staehely, "Intelligibility of modifications to dysarthric speech," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'03), Apr. 2003, vol. 1, pp. 924–927.

[23] T. Stephenson, H. Bourlard, S. Bengio, and A. Morris, "Automatic speech recognition using dynamic Bayesian networks with both acoustic and articulatory variables," IDIAP, Tech. Rep. 00-19, 2000

[24] K. Livescu, J. Glass, and J. Bilmes, "Hidden feature models for speech recognition using dynamic Bayesian networks," in Proc. of Euro speech '03, Geneva, 2003, pp. 2529–2532.

[25] K. Livescu, J. Glass, and J. Bilmes, "Hidden feature models for speech recognition using dynamic Bayesian networks," in Proc. of Euro speech '03, Geneva, 2003, pp. 2529–2532.