Accent Recognition for Indian English using Acoustic Feature Approach

Santosh Gaikwad
Research Fellow
Department of CS & IT
Dr.Babasaheb Ambedkar Marathwada University,Aurangabad.

Bharti Gawali
Associate Professor
Department of CS & IT
Dr.Babasaheb Ambedkar Marathwada University, Aurangabad.

K.V.Kale
Professor
Department of CS & IT
Dr.Babasaheb Ambedkar Marathwada University,Aurangabad.

ABSTRACT

Accent is the basic pattern of acoustic feature and pronunciation. It can identify the person's social and linguistic background. It is an important source of inter as well as intra speaker variability. The accent dependent dictionary or model can be used to improve accuracy of speech recognition system. In this study we present an experimental approach of acoustic speech feature for Marathi & Arabic accents for English speaking. The detail study of acoustics correlates the accent using formant frequency, energy and pitch characteristics. The database consists of speech from speaker with Marathi as their mother tongue and speakers from Iraq with Arabic language as mother tongue. Both the speakers were asked to speak English number from zero to nine. Through experimental results the fifth formant frequency found to be very effective for accent recognition.

Keywords

Accent, Acoustic ,Energy, Formant Frequency, Pitch, Foreign

1. INTRODUCTION

In the current era of research, there have been significance advances in speech recognition. There is wide range of acoustic feature containing speech signal that provide information about speaker background such as gender, accent, stress and emotion level [1]. If basic knowledge of speaker accent is estimated accurately then modified set of classification and recognition model could be employed to increase a recognition performance [2].

Accent is a pattern of pronunciation and acoustic feature which differentiate individual speech belonging to particular language group. Normally speakers, who speak language other than their own mother tongue, assign a non-voluntary appearance in their speech pattern. The age factor also plays an important role in accent identification, if speaker acquire knowledge of different language at an early age, his ability improves for the accent of particular language [3]. Every individual speaker develops characteristics of speaking style at an early age which will depend on native spoken language; therefore many features of native language persisted in speech.

Accent is a one of the second most important factor after gender that reduces accuracy of speaker dependant Recognition system [3, 4]. Most recognition systems are gender dependant there is no accent specific information utilized, which is the need of the current research. In order to motivate the problem of accent specific information retrieval and classification, we attempt to determine the power of accent on speech recognition performance.

The recognition and classification of accent is also challenging problem in speech recognition research .The recognition rate of French accented speaker of English was lower than that for native English speaker. This study has also been conducted which attempted normalize the command of regional accent prior to speech recognition task [5]. The British English accent normalization is carried by Barry et.al. Where the vowel quality difference was studied [6]. Speech prosodic feature for recognition was investigated by Weibel which results in improvement of speech recognition performance with the help of combining two sources of information [7]. Accent variation does not only stretch out in phonetic characteristics but also in prosodic characteristics [8] is shown by J.C.Wells. Group of people with similar geographical, linguistic, social and cultural background can be consider to share various common acoustic features resulting the similarity in accent as well as talking style, stress, tempo, all of which contribute speaker accent [9].Arslan and Hansan developed an English accent classification system which is most important and remarkable research in this area. The classification and performance increases as the count of test word increase [10]. This reference shows that accent recognition ability can improve the recognition of system.

The intensity has a strongest effect for differentiating between two accents based on F2, F3, and Vowel duration [11]. The F2, F3 and MFCC feature plays an important role in accent identification but inventive approach towards MLP with classification of SVM and KNN, which gives beneficial accuracy, is worked in [12]. The hybrid combination of frequency gives beneficial result towards accent recognition. [13].
The paper is structured in five sections. The database creation and preprocessing describe in section II. A basic acoustic feature is explained in section III. Section IV dedicated with experimental analysis followed by conclusion.

2. DATABASE CREATION

The corpus of accented speech was collected from 20 male and female subjects in two different groups, Arabic (speaker Number=10) and Indian Marathi (Speaker Number=10) .The corpus consist of annotated speech recorded over desktop computer with the help of close talk head mounted microphone. The database created for isolated digit (zero to nine). The speaker spontaneous response include gender, age, mother tongue, place of residence, primary schooling. Utterances were transcribed phonetically. The age of the speakers was between 20 and 35. All subjects were from department of CS & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad. The recording was done in morning and evening session in same location in order to minimize channel effects. The technical specifications of concentrated parameter for database creation are explained in table 1. The detailed English transcription of isolated word with symbol used for Arabic and Marathi accent is described in table 2.

Table 1: Technical specifications of concentrated parameter for database creation

Sr.No	Parameter	Specification
1	Sampling Frequency	16 KHz
2	Distance from microphone	10 cm
3	Environment	Office
4	Temperature	36.5 degree
5	Channel	Single
6	Gender	Male:15 Female:05
7	Total Vocabulary	2000 sample

The following formula was used for the nomenclature and labeling of database

$$\text{Labeled data} = \sum_{i=1}^{n} Ak$$

Where A is used for number of speaker and k is used for utterances index. We used the Praat and matlab tool for preprocessing. The data point was taken from 20 ms clips of utterances and was averaged over a window of 3 second to form feature. Various preprocessing technique was attempted including sliding window, hamming window, standardization and zero padding from data point.

3. ACOUSTIC FEATURES

It is noted in literature that F2-F3 contours, Pitch, Second and third formant frequency are effective acoustic features and found to be improving factors for performance. It improves the performance of speaker recognition system.[14,15, 16,17,18]

Table 2: English transcription of isolated word with symbol used for Marathi and Arabic accent

Sr.No	Isolated word	English transcription	Symbol used in this paper	
			Marathi accent	Arabic accent
1	ZERO	Z IY R OW	MA	AA
2	ONE	W AH N	MB	AB
3	TWO	T UW	MC	AC
4	THREE	TH R IY	MD	AD
5	FOUR	F AO R	ME	AE
6	FIVE	F AY V	MF	AF
7	SIX	S IH K S	MG	AG
8	SEVEN	S EH V AH N	MH	AH
9	EIGHT	EY T	MI	AI
10	NINE	N AY N	MJ	AJ

3.1 FREQUENCY CHARACTERISTICS

For the formulation of best speech recognition implementation technique it is beneficial to first consider aspects of human auditory perception parameter. Psychoacoustic analysis of human auditory perception mechanism shows that human ear responds differently to each acoustic tone based on its acoustic background. Frequencies characteristics differentiate gender, age group, as well as the accent [19]. Empirical evidence suggests that frequency play an important role in speech analysis. After extensive experimental analysis feature extraction technique formulated for the sampling of frequency criteria [20]. Table 3 below shows the standard frequency range of different age group. The production of the frequency from audio production model with of low and high frequency is described in figure 1.

Table 3: The standard frequency range of different age group

Sr.No	Age group(range)	Frequency (Hz)
1	20-29	119.5
2	30-39	112.2
3	40-49	107.1
4	50-59	118.4
5	60-69	112.2
6	70-79	132.2

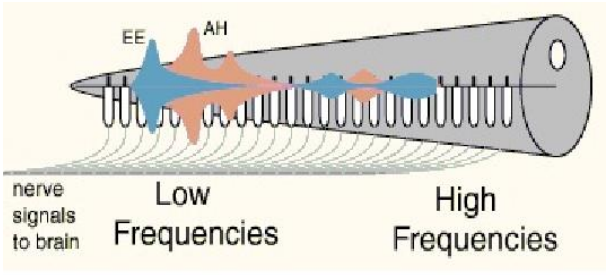

Figure 1: Production of frequency range from audio production system.

The formant frequency is very important for determining the phonetic background of speech. Normally non- native speaker's changes articulator moment in their own language when they speak language except their mother tongue. The experiment performed for investigation of influence of position of tongue constriction as well as constriction area of format frequency [16]. Several researchers contributed towards investigation of formant frequencies as a basic speech recognition features, using various methods for basic analysis [23, 24], synthesis with Fourier spectra [25], and picking on cepstral smoothed spectra [26].

3.2 PITCH CHARACTERISTICS

The pitch is a robust and dynamic feature for gender specific information recognition. The problem of pitch estimation has been addressed for a long time using many different approaches. In recent years, techniques like statistical learning [27], time domain probabilistic approaches for waveform analysis [28], or optimization techniques [29] have been applied to accomplish this task. However, most of these techniques are not robust enough, especially for corrupted speech.

3.3 ENERGY CHARACTERISTICS

The harmonic structure of speech is one of the most salient features. During the production of voice segment regular excitation of the vocal tract produces basically fundamental energy (F0) and it multiple combination. In the tonal language of speech allows expressing an emotion as well as accent information of speaker background. [15]. The energy is a robust and dynamic feature that can allow differences of speaking style of speaker and language information.

4. EXPERIMENTAL ANALYSIS

The experiment is performed for energy, pitch and formant frequency characteristics acoustic features.

4.1 Preprocessing

The preprocessing of dataset has been carried out using CSL and Praat software. In preprocessing parameter of speech signal such as spectrogram, energy contour, frequency contour, FFT waterfall was considered .The voiced period of speech signal was identified and applied to all experiment used for preprocessing of speech signal. The detail FFT waterfall model and spectrogram model of isolated word MB described in figure 2 and 3 respectively.

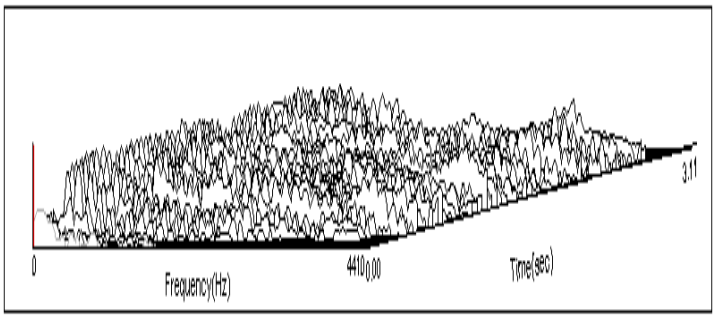

Figure 2: The FFT waterfall model of isolated word MB.

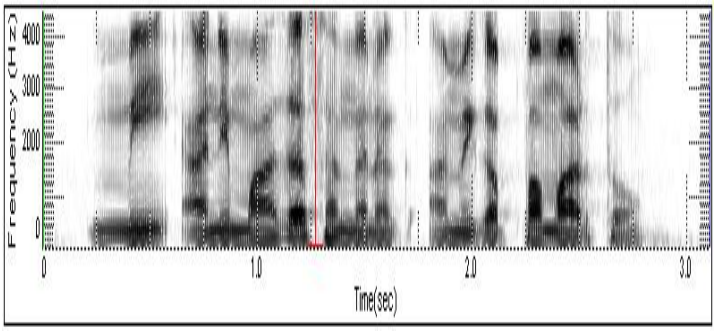

Figure 3: spectrogram variation of isolated word MB

4.2 Energy characteristics

The accent specific information recognition concentrate towards basic energy F0 and first as well as second level difference between that energy with average energy values. The F0, F1, F2 and average energy values specified efficiency about accent recognition for Marathi and Arabic is described in table 4.

The performance of accent recognition using different accent group as well as randomly selected mixed accent group with different training and testing environment is described in table 5. For independent accent group the training is of 600 samples and testing is done on 400 samples but for mixed dataset training environment is of 1000 sample and testing done for 500 samples. The average energy is standard flow towards speaking style. The graphical representation of average energy for MA and AA are presented in figure 4.

Table 4: Specified efficiency of average energy values

Sr.No	Name of speech	Average Energy Values		
		Arabic	**Marathi**	**Difference**
1	Zero	3.174	0.3079	2.8661
2	One	1.3197	0.2193	1.1004
3	Two	2.4882	0.2435	2.2447
4	Three	0.8427	0.1757	0.667
5	Four	2.1166	0.267	1.8496
6	Five	1.382	0.1974	1.1846
7	Six	0.6832	0.0314	0.6518
8	Seven	1.2153	0.1321	1.0832
9	Eight	1.085	0.1073	0.9777
10	Nine	0.8626	0.1775	0.6851

Table 5: The performance of accent recognition for different training and testing environment.

Sr.No	Accent Group	Energy Feature	Training Dataset	Testing Dataset	Accuracy (%)
1	Speaker with Marathi accent	Fo	600	400	92.18
		F1	600	400	89.09
		F2	600	400	87.90
2	Speaker with Arabic Accent	Fo	600	400	93.21
		F1	600	400	93.02
		F2	600	400	92.76
3	Test Dataset	Fo	1000	500	90.67
		F1	1000	500	89.90
		F2	1000	500	89.05

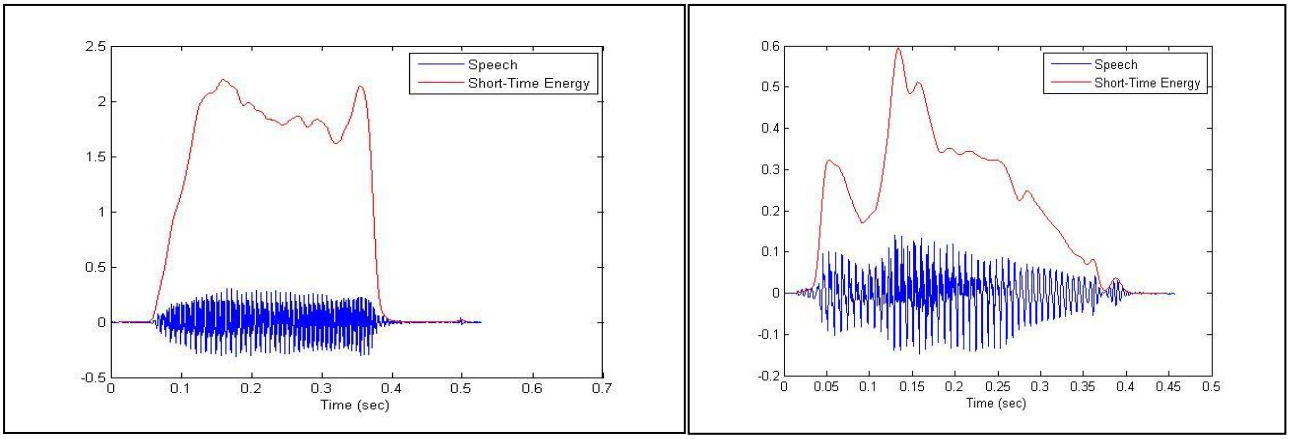

Figure 4: Graphical representation of average energy for MA and AA samples.

4.3 Pitch Characteristics

For this pitch based accent identification, we used the mean, standard deviation and covariance of the Marathi and Arabic accent. The performance of pitch accent recognition is tabled in table 6.

Table 6: The performance of accent recognition using the pitch feature.

Sr.No	Accent Group	Pitch	Training vocabulary size	Testing vocabulary size	Accuracy (%)
1	Speaker with Marathi accent	Mean	600	400	87.40
		STD	600	400	84.23
		Variance	600	400	85.12
		covariance	600	400	90.80
2	Speaker with Arabic Accent	Mean	500	450	84.78
		STD	500	450	84.10
		Variance	500	450	86.40
		covariance	500	450	89.91
3	Test Dataset	Mean	1000	500	81.89
		STD	1000	500	80.09
		Variance	1000	500	79.12
		covariance	1000	500	86.34

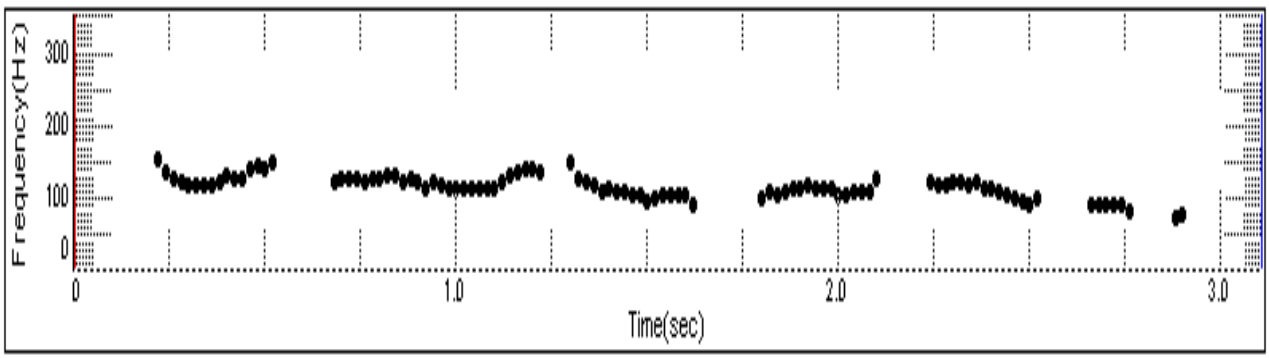

Figure 5: Graphical representation of pitch feature of MB sample.

4.4 Formant Frequency

The performance of the formant based accent recognition tested on different training and testing environment with second, third, fourth and fifth formant frequency is shown in table 7. This experiment focuses on fifth formant frequency values. The fifth formant frequency values give an efficient result. The graphical representation of formant frequency of same voice sample uttered by different Marathi and Arabic accent figured in 5 and 6.

Table 7: The performance of accent recognition using formant frequency.

Sr.No	Accent Group	Formant Level	Training vocabulary size	Testing vocabulary size	Accuracy	Confused	Error rate
1	Speaker with Marathi accent	Second	600	400	91.38	4.2	4.42
		Third	600	400	92.34	4.6	3.06
		Forth	600	400	88.90	5.7	5.4
		Fifth	600	400	93.28	3.4	3.32
2	Speaker with Arabic Accent	Second	500	450	86.78	5.8	7.42
		Third	500	450	86.18	4.9	7.92
		Forth	500	450	83.24	5.09	11.67
		Fifth	500	450	91.10	6.2	2.7
3	Test Dataset	Second	1000	500	83.75	6.90	9.35
		Third	1000	500	84.02	6.78	9.02
		Forth	1000	500	78.56	7.19	17.25
		Fifth	1000	500	88.17	8.02	3.81

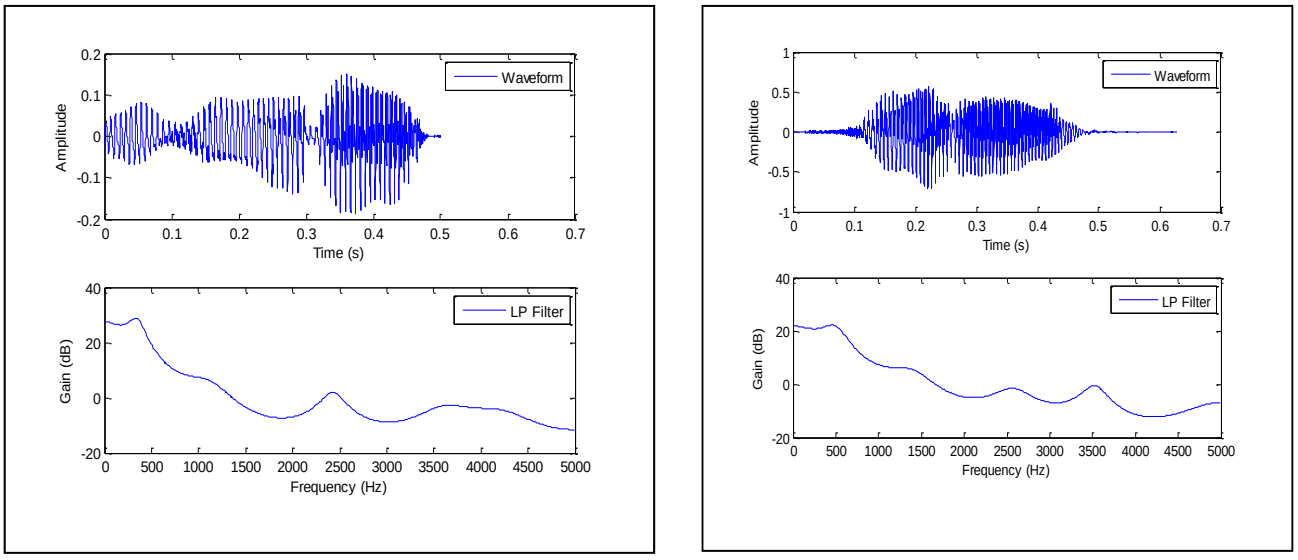

Figure 6: A Graphical representation of formant frequency for MA and AA samples.

The comparative performance of accent recognition on the basic of energy, pitch and formant frequency is described in table 8.

Table 8: comparative performance of accent recognition on the basis of energy, pitch and formant frequency

Sr.No	Accent Group	Acoustic feature	Training vocabulary size	Testing vocabulary size	Accuracy (%)
1	Speaker with Marathi accent	Energy	600	400	92.18
		Format Frequency	600	400	93.28
		Pitch	600	400	90.80
2	Speaker with Arabic accent	Energy	500	450	93.21
		Format Frequency	500	450	91.10
		Pitch	500	450	89.91
3	Test Dataset	Energy	1000	500	90.67
		Format Frequency	1000	500	88.17
		Pitch	1000	500	86.34

The experiment resulted in following observation:

1. In the accent specific information retrieval, the basic energy acoustic features were analyzed. In this study, we approached towards difference between the energy. The average energy values of Arabic accent found to be higher than Marathi accent.

2. F0 Basic level energy is high in Marathi and Arabic accent dataset whereas in the mixed dataset the F1 and F2 were found higher.

3. We observed that fifth formant frequency was also effective for accent recognition.

4. This study attempted the pitch feature of speech for accent recognition which resulted in efficient accent recognition.

5. CONCLUSION

This work presents the individual as well as comparative analysis of acoustic feature such as energy, formant frequency and pitch for Arabic and Marathi accented Indian English. The result of accent recognition analysis shows that average energy, formant frequency and pitch feature varies in Marathi and Arabic accent. When the system is compiled independently for Marathi accent group the formant frequency gives effective performance than other acoustic feature. The energy found to be high in Arabic accent group, than other acoustic features. The experiment tested on dataset that includes database of Marathi accent as well Arabic accent shows that not only F2, F3 but F5 can also be effectively used for accent recognition research.

6. ACKNOWLEDGEMENT

The authors wish to acknowledge the DST under Fast Track Scheme entitled as “Design and Development of Marathi Speech Interface System”. And financial support of UGC in Special Assistance Program (SAP) DRS Phase-I for research work on the theme “Biometrics: Multimodal System Development”.

7. REFERENCES

[1] A. Ikeno and J.H.L. Hansen, "The effect of listener accent background on accent perception and comprehension", EURASIP Journal on Audio, Speech, and Music Processing, Vol. 2007, Article ID 76030, 8 pages, 2007.

[2] C. Pedersen and J. Diederich, "Accent classification using support vector machines", 6th IEEE/ACIS International Conference on Computer and Information Science , ICIS 2007.

[3] Asher and G. Garcia (19691, “The optimal age to learn a foreign language”, Modern Language J., Vol. 38, pp. 334- 341.

[4] L.R.Rabiner and J.G Wilpon (1977),”Speaker Independent isolated word recognition for a moderate size (54 word) vocublary”, IEEE Trans.Acoust.Speech signal processing Vol.27,pp.538-587.

[5] V. Gupta and P. Mermelstein (1982). “Effects of speaker accent on the performance of a speaker-independent, isolated-word recognizer”, J. Acoust. Sot. Amer., Vol. 71, pp. 1581- 1587.

[6] W.J. Barry, C.E. Hoequist and F.J. Nolan (19891, “An approach to the problem of regional accent in automatic speech recognition”, Computer Speech and Language, Vol. 3, pp. 355-366.

[7] A. Ljolje and F. Fallside (1987), “Recognition of isolated prosodic patterns using hidden Markov models”, Computer Speech and Language, Vol. 2, pp. 27-33.

[8] J.C. Wells, Accents of English, volume:1,2, Cambridge University Press, 1982.

[9] C. Pedersen and J. Diederich, "accent classification using support vector machines", 6th IEEE/ACIS International Conference on Computer and Information Science , ICIS 2007.

[10] L. M. Arslan and J.H.L. Hansen, "Language accent classification in American English", Speech Communication, Revised January 29, 1996.

[11] Zheng, D. C., Dyke, D., Berryman, F., & Morgan, C. (2011). A new approach to acoustic analysis of two British regional accents—Birmingham and Liverpool accents. International Journal of Speech Technology, 15(2), 77–85. doi:10.1007/s10772-011-9123-3.

[12] Rabiee, A., & Setayeshi, S. (2010). Persian Accents Identification Using an Adaptive Neural Network. 2010 Second International Workshop on Education Technology and Computer Science, 7–10. doi:10.1109/ETCS.2010.273

[13] Hansen, J. H. L., Arslan, L. M., & Carolina, N. (1997). Frequency characteristics of foreign accented speech, Duke University Department of Electrical Engineering, 1123–1126.

[14] Tang, H., & Ghorbani, A. A. (n.d.). Accent Classification Using Support Vector Machine and Hidden Markov Model, (1), 3–4.

[15] Arslan, L. M., & Hansen, J. H. L. (1996). Language accent classification in American English. Speech

Communication, 18(4), 353–367. doi:10.1016/0167-6393(96)00024-6 .

[16] Hansen, J. H. L., Arslan, L. M., & Carolina, N. (1997). Frequency characteristics for foreign accented speech, Duke University Department of Electrical Engineering, 1123–1126.

[17] Kat, L. I. U. W., Fung, P., Bay, C. W., & Kong, H. (1999). Fats accent identification and accented speech recognition, 3–6.

[18] Xuejing Sun . Pitch accent prediction using ensemble machine learning, Department of Communication Sciences and Disorders, Northwestern University 2299 N. Campus Dr., Evanston, IL 60208, USA.

[19] Variation of vocal format and speech [online] http://hyperphysics.phy-astr.gsu.edu/hbase/music/vowel2.html viewed on 02 Sept 2012.

[20] Zissman, M. (1993). Automatic language identification using Gaussian mixture and hidden Markov models. Acoustics, Speech, and Signal Processing, 1993 …, 399–402. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=319323

[21] Hanani, A., Russell, M., & Carey, M. J. (2011). Speech-based identification of social groups in a single accent of British English by humans and computers. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4876–4879. doi:10.1109/ICASSP.2011.5947448

[22] Zheng, Y., Sproat, R., Gu, L., Shafran, I., Zhou, H., Su, Y., Jurafsky, D., et al. (n.d.). Accent Detection and Speech Recognition for Shanghai-Accented Mandarin, 7–10.

[23] John N. Holmes, Wendy J. Holmes and Philip N. Garner "Using formant frequencies in speech recognition, Speech Technology Consultant, 19 Maylands Drive, Uxbridge, UB8 1BH, U.K.

[24] P. Schmid and E. Barnard, "Robust, N-Best Formant Tracking", Proc. EUROSPEECH'95, pp. 737-740, Madrid, 1995

[25] L. Welling and H. Ney, "A Model for Efficient Formant Estimation", Proc. IEEE ICASSP, pp. 797-800, Atlanta, 1996

[26] Y. Laprie and M.-O. Berger, "Active Models for Regularizing Formant Trajectories", Proc. ICSLP, pp. 815-818, Banff, 1992

[27] Levow, G. (2009). Investigating Pitch Accent Recognition in Non-native Speech, (August), 269–272.

[28] Ishi, C. T., Hirose, K., & Minematsu, N. (2003). Mora F0 representation for accent type identification in continuous speech and considerations on its relation with perceived pitch values. *Speech Communication*, *41*(2-3), 441–453. doi:10.1016/S0167-6393(03)0014-1

[29] Stantic, D., & Jo, J. (2012). Accent Identification by Clustering and Scoring Formants, 232–237.