A Meter Classification System for Spoken Persian Poetries

Saeid Hamidi saeidhamidi@iau-maragheh.ac.ir
Department of Electrical Engineering,
Maragheh Branch
Islamic Azad University
Maragheh,Iran

Farbod Razzazi razzazi@srbiau.ac.ir
Department of Electrical and Computer Engineering,
Science and Research Branch
Islamic Azad University
Tehran,Iran

Abstract

In this article, a meter classification system has been proposed for Persian poems based on features that are extracted from uttered poem. In the first stage, the utterance has been segmented into syllables using three features, pitch frequency and modified energy of each frame of the utterance and its temporal variations. In the second stage, each syllable is classified into long syllable and short syllable classes which is a historically convenient categorization in Persian literature. In this stage, the classifier is an SVM classifier with radial basis function kernel. The employed features are the syllable temporal duration, zero crossing rate and PARCOR coefficients of each syllable. The sequence of extracted syllables classes is then softly compared with classic Persian meter styles using dynamic time warping, to make the system robust against syllables insertion, deletion or classification and poems authorities. The system has been evaluated on 136 poetries utterances from 12 Persian meter styles gathered from 8 speakers, using k-fold evaluation strategy. The results show 91% accuracy in three top meter style choices of the system.

Keywords: Syllable Classification, Utterance Syllabification, Automatic Meter Detection, Support Vector Machines, Dynamic Time Warping, Poetries Categorization

1. INTRODUCTION

Poem is the vital part of literature of all cultures and reflects the specifications and maturity of a cultural society. Rhyme and meter are considered as inseparable elements of the poetry and meter extraction is a historically exquisite subject for literary scholars which have been extracted intuitively by now.

There is a rich literature on automatic speech recognition systems for general applications in last 30 years; however, automatic extraction of rhyme and meter styles from uttered poems is the focus of some studies in recent years for various languages including Chinese [1-3], Thai [4] and European languages [5,6]. Although there is a very rich treasury of Persian poetries which are created during more than are thousand years, however, most of the studies on these poems are literary studies and they are not well prepared for machinery manipulations. In particular, there is little research, concentrating on machinery Persian meter detection in uttered speech.

In this article, automatic speech recognition utilities are employed to extract an algorithm for automatic meter detection from uttered poetries. The input of this system is a single uttered verse of a poem and the output is the meter style. The organization of the paper is as follows. The theory of Meter extraction in Persian poetries is introduced in section 2. In section 3, the architecture of the proposed system is discussed. The syllabification algorithm is presented in section 3. Section 4 and 5 is devoted to syllable classification and sequence classification of syllables respectively. In section 6, the implementation of the system is analyzed and evaluated. The paper is concluded in section 7.

2. THEORY OF METER DETECTION IN PERSIAN POETRIES

There are a few studies on extracting and detecting the poetries meter and rhyme in different languages [1-5]. However, the poetry is a specific property of each language and the meter extraction problem should be handled separately in each language.

Thank to the theory of Persian meters, named as Arooz, with more than 700 years old, Persian meter detection is based on syllabification of speech into long and short syllables [7,8]. There is a set of distinguished classes of Persian meter styles in the literature which 12 classes of them covers most of the existent poetries.

The categorization of Persian syllables is tabulated in TABLE 1. As demonstrated, the variety of Persian syllables is limited. There is a vowel in the kernel of each syllable and there is one starting consonant before the vowel. After the vowel, it is possible to have no consonant, one consonant or two consonants. This simple structure motivates us to find out the syllabification sequence by extracting the location of kernel vowel and locating the boundaries of the syllables by moving front and back from the kernel.

#	Syllable Category
1	CV
2	CVC
3	CVCC

TABLE 1: Variety of Persian Syllables

In Arooz Theory, there are two kinds of syllables, short syllables and long syllables. Long and short syllables are distinguished by the kernel vowel used in the syllables. The vowel in short syllables are a member of the set (/ae/,/eh/,/oy/) . In contrast, long syllables consist a vowel in the set (/aa/,/iy/,/ux/) The Nomination short and long syllables is due to the utterance duration of each syllable when it is intended to read the poem in meter style. It is empirically shown that the times spent to utter long syllables are almost similar. Short syllables are uttered in similar duration too. However, the duration of short and long syllables are not the same.

It is revealed that there are common standard meters that are frequently used by poets and seems to be well accepted by Persian speakers. Over 95% of the poetries are covered by 12 standard meters. Therefore, this study is concentrated on these standard meters which are tabulated in TABLE 2.

All of the verses of each poem should employ similar standard meter. However, there is no strict rule in the artistic world. Although, most of the verses employ one standard meter, in fact, sometimes, poets have used the standard meter by slight modifications in some verses. This phenomenon, which is called poetry authorities, will make the approach of the machinery system to be a soft likelihood measurement rather than pattern matching.

#	Standard Meter	Example written in Phonetics	Example written in Persian
1	SSLS LSLL SSLS LSLL	q/ae/l/iy/eh/y/hh/oy/m/aa/y/eh/r/ae/hh/m/ae/t/ oy/ch/eh/aa/y/ae/t/iy/kh/oy/d/aa/r/aa	علی ای همای رحمت تو چه آیتی خدا را
2	LSSL SLSL LSSL SLSL	s/ae/r/v/eh/ch/ae/m/aa/n/ae/m/ae/n/ch/eh/r/aa/ m/eh/y/l/eh/ch/ae/m/ae/n/n/ae/m/iy/k/oy/na/d	سرو چمان من چرا میل چمن نمیکند
3	SSLL SLSL SSL	d/ae/r/d/eh/q/eh/sh/gh/iy/k/ae/sh/iy/d/eh/q/ae/m/ k/eh/m/ae/p/oy/r/s	درد عشقی کشیده ام که مپرس
4	SLSL SSLL SLSL SSL	t/ae/n/ae/t/b/eh/n/aa/z/eh/t/ae/b/iy/b/aa/n/ n/iy/y/aa/z/m/ae/n/d/m/ae/b/aa/d	تنت به ناز طبیبان نیازمند مباد
5	SLL SLL SLL SL	m/ae/g/ae/r/d/aa/n/s/ae/r/ae/z/ d/iy/n/oy/q/ae/z/r/aa/s/t/iy	مگردان سر از دین و از راستی
6	SLL SLL SLL SLL	n/eh/k/ou/h/eh/sh/m/ae/k/oy/n/ ch/ae/r/kh/eh/n/iy/l/oy/f/ae/r/iy/r/aa	نکوهش مکن چرخ نیلوفری را
7	LLSL LLSL LLSL LLSL	b/aa/m/ae/n/b/eh/g/oy/t/aa/k/iy/s/t/iy/ m/eh/h/r/iy/b/eh/g/ou/m/aa/h/iy/b/eh/g/ou	با من بگو تا کیستی مهری بگو ماهی بگو
8	SLLL SLLL SLLL SLLL	b/iy/aa/t/aa/g/oy/l/b/ae/r/ae/f/sh/aa/n/ iy/m/oy/m/eh/iy/d/ae/r/s/aa/gh/ae/r/ ae/n/d/aa/z/iy/m	بیا تا گل برافشانیم و می در ساغر اندازیم
9	LLSL SLL LLSL SLL	y/aa/r/ae/b/t/oy/q/aa/sh/ae/n/aa/r/aa/ m/oy/h/l/ae/t/d/eh/h/oy/s/ae/l/aa/m/ae/t	یارب تو آشنا را مهلت ده و سلامت
10	LLS LLL LLS SLL	q/ae/iy/p/aa/d/ae/sh/ae/h/eh/kh/ou/b/aa/ n/d/aa/d/ae/z/gh/ae/m/eh/t/ae/n/h/aa/q/iy	ای پادشه خوبان داد از غم تنهائی
11	LSLL LSLL LSL	m/ae/r/d/eh/r/aa/d/ae/r/d/iy/ae/g/ae/r/ b/aa/sh/ae/d/kh/oy/sh/ae/s/t	مرد را دردی اگر باشد خوشست
12	SSLL SSLL SSLL SSLL	d/ae/r/d/eh/l/ae/m/b/ou/d/k/eh/b/iy/d/ou/s/t/ n/ae/b/aa/sh/ae/m/h/ae/g/ae/z	در دلم بود که بی دوست نباشم هرگز

TABLE 2: Categories of frequently used Persian standard meters

3. SYSTEM ARCHITECTURE

FIGURE.1 presents the overall block diagram of meter extraction for Persian poetries. As it is shown, after preprocessing and syllabification, some features are extracted from each syllable. The features are zero-crossing rate, PARCOR coefficients and temporal duration for each syllable.

In the next stage, syllables are classified into long and short syllables. Finally the sequence of syllable classes is compared to standard Persian poetry meters using dynamic programming. The best match of the sequence with the standard meters shows the category of the meter.

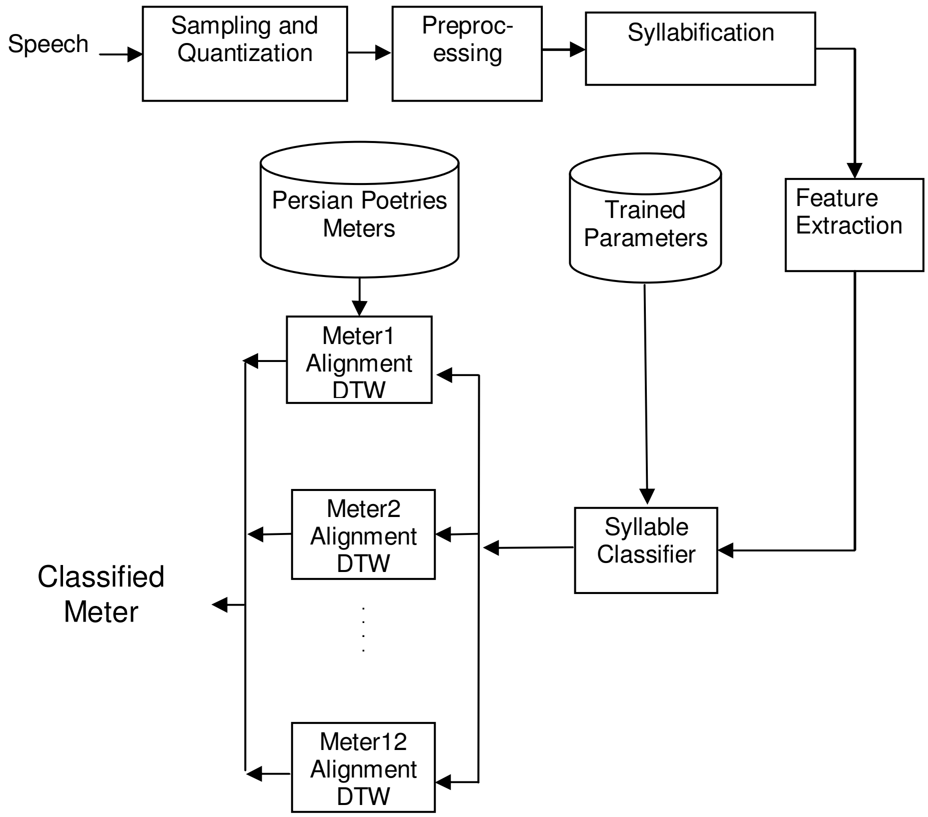

FIGURE 1: Overall Block Diagram

4. SYLLABIFICATION SUBSYSTEM

The meters in Persian poetries are categorized based on the sequence of syllables classes. Therefore, it is essential to have a syllabification stage [9-11]. Syllables in Persian consist of one consonant, one vowel and probably one or two consonants respectively. As a result, syllable segmentation is based on detecting the location of vowels in the utterance. The implemented syllable segmentation is based on three features, the pitch frequency, the energy and estimation of energy derivative during time.

FIGURE 2 demonstrates the detail of mentioned syllable segmentation procedure. After a 32ms, 50% overlapped framing, the utterance is filtered to have low pass frequency components and detect the pitch frequency more accurately. The energy and pitch frequency of the frame are then extracted. These two features are used to extract the boundary frames of syllables.

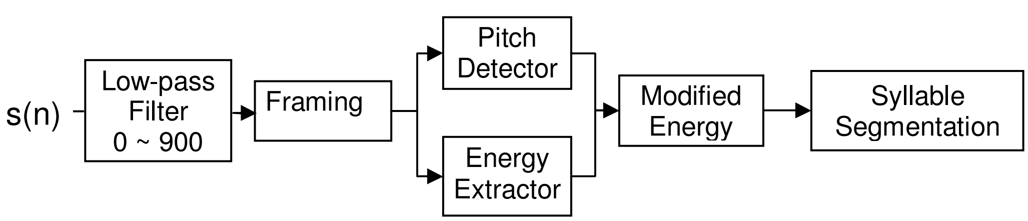

FIGURE 2: Syllable Segmentation Block Diagram

The segmentation algorithm requires locating the voiced frames. Hence, the focus of the pitch detection algorithm is the extraction of the frames where pitch frequency can be recovered in them. As demonstrated in FIGURE 3, the voiced and unvoiced frames of the utterance are segmented based on Dubnowski - Rabiner algorithm [12].

To extract the modified energy, the energy of the frame is computed as the sum of squared amplitude of frame samples.

$$E[n] = \sum_{i=1}^{L} x^2(i)$$

where E[n] is the nth frame energy and L is the length of the nth frame. x represents the frame samples amplitude. The frame energies are then normalized to maximum frame energy.

$$E_{norm}[n] = \frac{E[n]}{Max(E[n])}$$

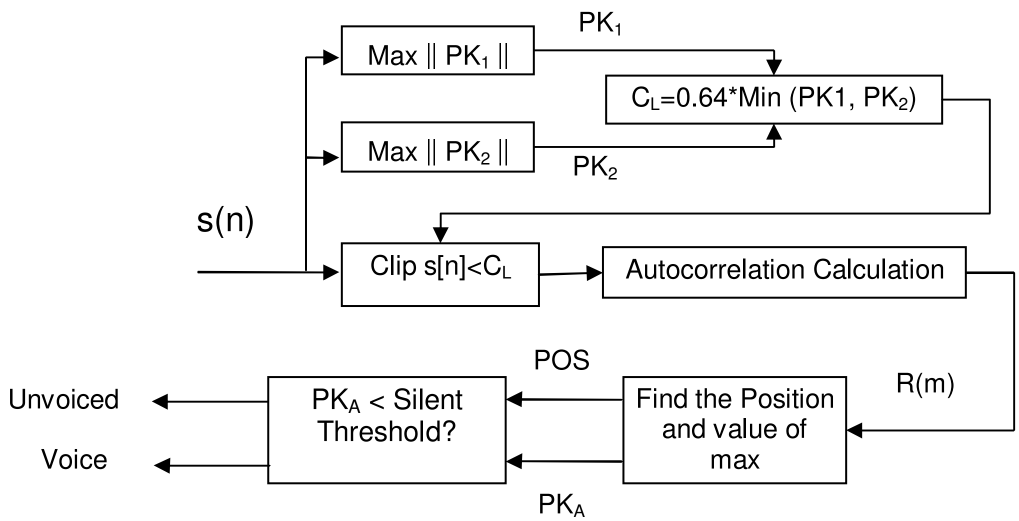

FIGURE 3: Voiced/ Unvoiced Classification

The energy contour of the utterance is then filtered by a nonlinear median filter to enhance the locating procedure of peaks and valleys in energy contour and therefore enhancing the performance of syllable segmentation stage. Modified energy is then extracted as the windowed version of smoothed energy by voiced framed locations.

FIGURE 4 demonstrates a sample of segmented signals into voiced and unvoiced segments. The modified energy is the clipped version of smoothed energy. Pitch and modified energy contours are nonzero in voiced frames of the utterance, while the smoothed energy is nonzero for all frames.
Modified Energy is the input of syllabification stage. As it is shown in FIGURE 5, in this stage, the first valley in nonzero modified energies is the primary estimation of syllable boundary. To extract the subtle boundary, it is better to trace back the signal to find out the first zero in modified energy contour. To avoid undesired over-segmentation, the value of modified energy in the valley should be less than two thresholds derived by two adjacent peaks.

The thresholds are set to suitable empirical values to avoid unwanted syllabifications, a minimum length for syllables was considered. The algorithm continues until the end of the utterance. FIGURE 6 is an example of the procedure output. In this example, the utterance is /t ae v aa n aa b oy v ae d/ and the segmentation process succeeded to segment the utterance into (/t/ae/, v/aa, /n/aa, /b/oy, /v/ae/d)(توانا بود).

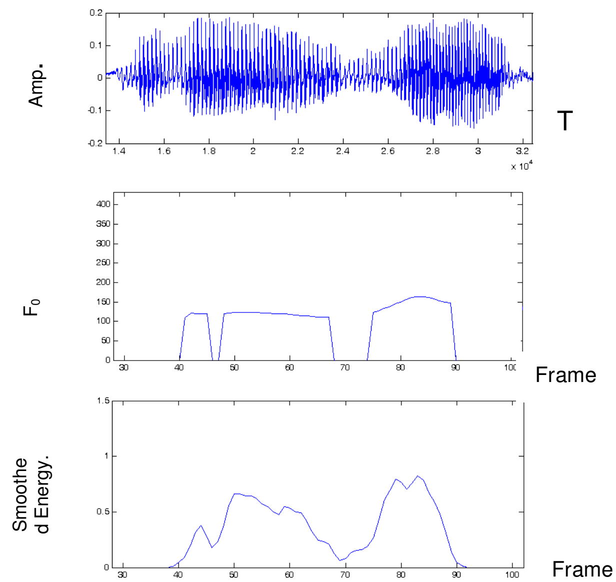

FIGURE 4: Sample of Intermediate Signals of Block Diagram in FIGURE 3 Top: input Signal Middle: Pitch Contour Bottom: Smoothed Energy.

5. SYLLABLE CLASSIFICATION

To classify the syllables into long and short syllables, 12 partial reflection coefficients are extracted from linear prediction analysis of each syllable in addition to zero crossing rate and syllable duration. These features are added up with two previous features extracted in the syllabification stage (pitch frequency, modified energy), make the whole feature vector. Intuitively, it seems that the most effective feature in syllable classification would be the syllable duration. Therefore the tests were designed to check the performance of the system on both single duration feature and the whole vector as the feature vector. Each feature is normalized with respect to mean and variance of the feature in the whole syllables space. Hence, all features become zero mean, unity variance after normalization.

The features are classified using a kernel based support vector machine with RBF kernel [13,14]. Kernel meta-parameters are optimized empirically using grid search [15] evaluated on K-fold evaluation strategy.

The overall block diagram is depicted in FIGURE 5. the modified energy is the input of syllable classification. The output of the block is both the start point and ending point of the syllable. The local maxima of the modified energy are the candidates of starting and ending point of the syllable. The candidates are refined using a minimum energy candidate. A sample of modified energy schematic behavior and real world behavior is demonstrated in FIGURES 6 and 7. The notations b and p demonstrate the central and boundary points of the syllable.

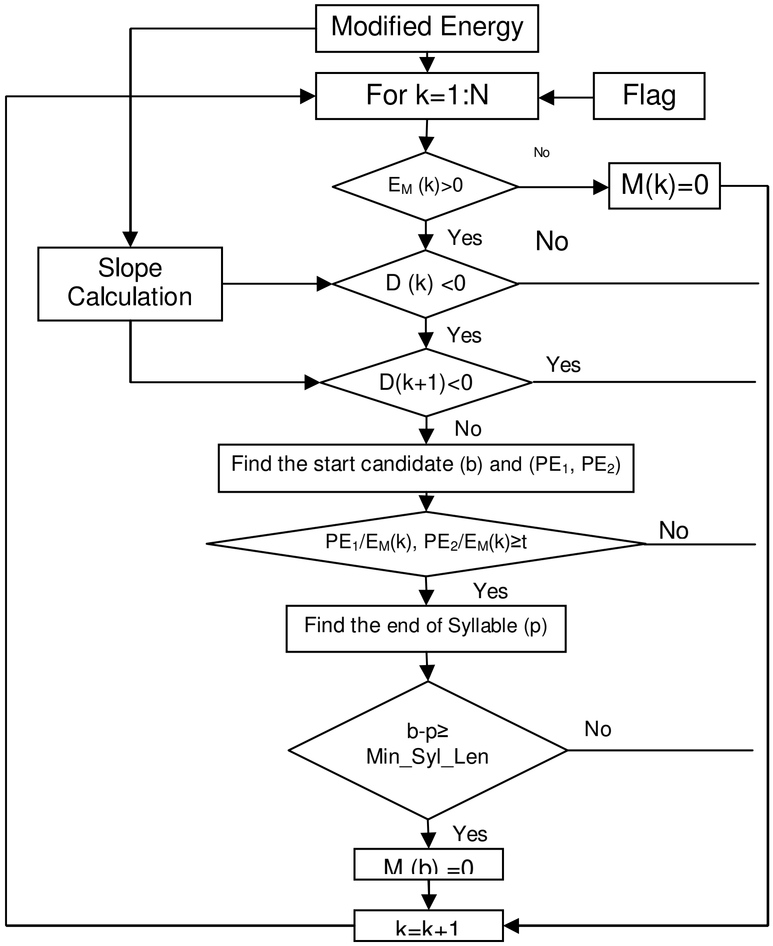

FIGURE 5: Syllabification Block Diagram

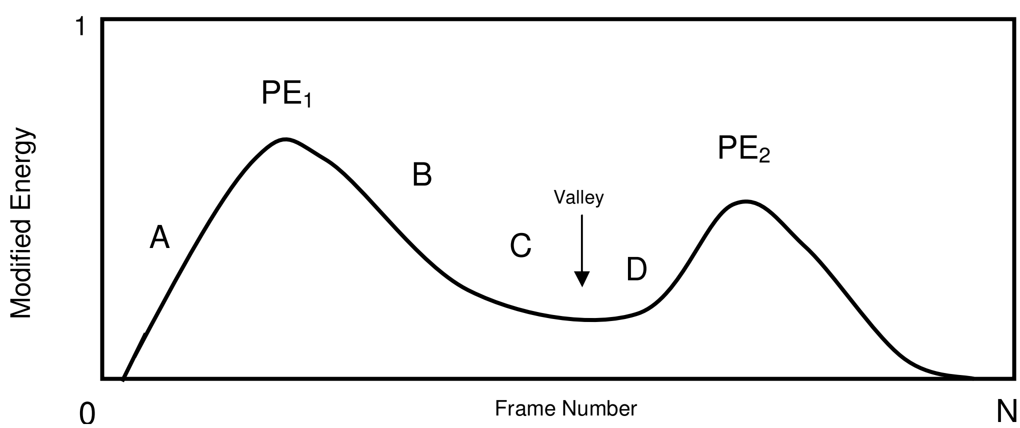

FIGURE 6: Modified Energy Variations in a Syllable

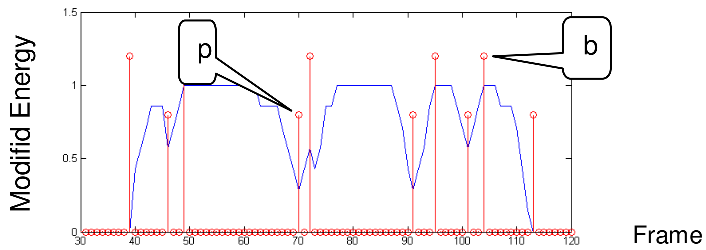

FIGURE 7: A Sample of Syllabification Based on Modified Energy

6. METER MATCHING USING DYNAMIC TIME WARPING

The meters are commonly described by a sequence of syllable classes. Fortunately, the employed meters in Persian poems are selected from a limited number of standard meters. This fact lets the meter classifier system to compensate syllable segmentation and classification errors. In addition, poets are not rigidly loyal to standard meters, however they use long syllables instead of short syllables and vise versa depending on linguistic and semantic constraints. This phenomenon, known as "poetry options" in the literature, should be considered to find out the best match (not exact match) between the extracted syllable classes sequence and the standard meters, including substitution, deletion and insertion errors. This matching was carried out by well known dynamic programming approach in speech utterances matching, named as dynamic time warping [16]. The output of the system is the number of matches, substitutions, insertions and deletions of one verse utterance with respect to each of standard meters. The error is defined as the sum of substituted, inserted and deleted syllables. The standard meter with minimum error is referred as the recognized meter.

7. EXPERIMENTS AND RESULTS

The proposed system was evaluated on 17 versus with 12 distinct Persian standard meters. The meters were selected to cover more than 95% of the Persian poetries. The verses were uttered by 8 native speakers whose are requested to pronounce each verse in correct meter. For evaluation purposes, all verses were manually segmented into syllables and all syllables were labeled by long and short syllable labels.

The first evaluation was made on the syllabification stage. The automatically extracted syllables were compared to manual segmentations and the number of insertions and deletions were evaluated. TABLE 3 explains the results in detail. As it can be observed, the average syllabification error is 10% which is comparable to the literature for other languages. The accuracy is variable with respect to the speaker, due to the pronunciation and accents variations. The standard deviation of this error in the set of speakers is about 19% of the mean value.
The system parameters are optimized to achieve the highest accuracy. In this point, the number of deletions is about twice the number of insertions.

In syllable classification stage, the system was evaluated by the classification rate. To optimize the classifier, two parameters (i.e. the misclassification weight in the training procedure (denoted as C) and the RBF kernel parameter denoted as □) were optimized in the logarithmic grid search basis.

The evaluation was performed based on K-Fold strategy with K=10. The average classification rates in 10 folds are tabulated in TABLE 4 for different C and γ values. It can be concluded that the system is optimized with (C,γ)=(3.16, 0.0316). The best syllable classification rate is about 75%. This accuracy rate may cause the sequence classification unusable unless the result is post-processed by a dynamic comparison with the reference meters.

Speaker	Number of Segments	Deletion	Insertion	Error Percentage
Spk1	433	0.1	0.01	11.7
Spk2	455	0.07	0.037	11.2
Spk3	434	0.108	0.02	12.9
Spk4	471	0.036	0.042	7.8
Spk5	448	0.078	0.024	10.2
Spk6	454	0.074	0.026	10.1
Spk7	459	0.074	0.045	11.9
Spk8	467	0.038	0.027	6.6
Ave.	452.6	32.8	13.7	10.2

TABLE 3: Syllable Segmentation Rates For Different Speakers

C	γ=0.01	γ=0.03	γ=0.1	γ=0.31
0.0001	62.66	62.66	62.66	62.66
0.00033	62.66	62.66	62.66	62.66
0.001	62.66	62.66	62.66	62.66
0.003	62.66	62.66	62.66	62.66
0.01	62.66	62.66	62.66	62.66
0.03	62.66	62.66	62.66	62.66
0.1	61.90	70.07	70.37	63.90
0.31	70.96	73.36	73.93	68.69
1	73.21	74.27	74.80	72.03
3.16	73.85	75.41	74.20	71.65
10	74.19	74.49	72.80	70.04
31.32	74.28	73.69	70.93	69.26
100	74.63	72.68	68.82	69.36
316.22	73.04	72.12	67.59	69.36

TABLE 4: SVM Classifier Meta-Parameter Optimization

In The last stage, the mentioned post-processing is carried out. A dynamic time warping algorithm was employed to compare the meters and select the best among common reference Persian poems meters. TABLE 5 and FIGURE 7 demonstrate the result for this dynamic matching and scoring.

The results in TABLE 5 show that the system achieves 91% in 3 best meter classification rate.

n-Best	Spk 1	Spk 2	Spk 3	Spk 4	Spk 5	Spk 6	Spk 7	Spk 8	Ave
1	64	64	58	64	76	76	76	70	69
2	88	94	76	94	88	100	82	94	89
3	88	100	76	94	94	100	82	94	91

TABLE 5: N-Best Classification Rates For Different Speakers

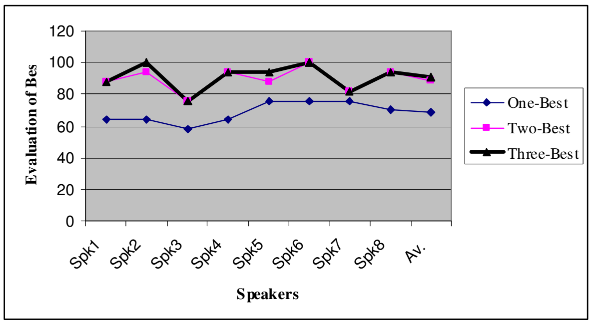

FIGURE 8: Comparison of n-best Rates for Different Speakers

The comparison of speakers in FIGURE 8 shows that there is a significant difference in the meter detection accuracy for different speakers. Probing the result, it was observed that bad results belong to the speakers who did not utter the poems in correct meter. Therefore, in this situation, incorrect detection of meter type is expectable. The system may be generalized to reject the miss-uttered poems by speakers.

8. CONSLUSION & FUTURE WORK

In this paper, an automatic meter detection algorithm was implemented and evaluated. This is the first attempt to analyze the utterance of Persian poetries automatically which may lead the researchers toward a new approach in investigating the literature theoretically and practically. In addition, there is a rich source of human culture in poetries, which can be digitized by tracing this trajectory.

9. REFERENCES

[1] Z. S. He, W.T. Liang, L. Y. Li, Y.F. Tian, *"SVM-Based Classification Method For Poetry Style"*, Proceedings of Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, Aug 2007, PP. 2936-2940.

[2] X.Wang, Y. Liu, L.Cai, *"Entering Tone Recognition in a Support Vector Machine Approach"*, Proceedings of Fourth International Conference on Natural Computation, Aug. 2007, PP. 61-65.

[3] Y.YI, Z. S. He, L. Y. Li, T. Yi, E. Yi, *"Advanced Studies On Traditional Chinese Poetry Style Identification"*, Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhu, Aug. 2005, PP. 3830-3833.

[4] N. Sataravaha. *"Tone classification of syllable segmentation Thai speech based on multilayer perceptron"* Acoustical Society of America Journal, Volume 111, Issue 5, 2002,pp. 2366-2366.

[5] H.R. Tizhoosh, R.A. Dara, *"On Poem Recognition"*, Pattern Analysis and Application, Vol. 9, No.4, Nov. 2006, PP. 325-338.

[6] H. Meinedo *"The use of syllable segmentation information in continuous speech recognition hybrid systems "* Applied to the Portuguese language , IEEE Transaction on Acoustic , speech and signal processing. 0-1623-8, 2000

[7] Vahidian-Kamkar T, *"Persian Poetries Meter and Rhyme"*, Nashr-e-Daneshgahi, 1990. (Published in Persian)

[8] Shamisa S, *"An Introduction to Meter and Rhyme"*, Ferdos Publications, 1992, (Published in Persian)

[9] Liany-Yan Li *, " Poetry stylistic analysis technique based on term connection "* . IEEE Speech and signal processing. 0-7803-8403, 2004

[10] Hartmut R.Pfitzinger *" syllable detection in read and spontaneous speech "* 13-4822-13324, 1996

[11] I. Kopecek: *"Speech recognition and syllable segmentation"*; in Proceedings of the Workshop on Text, Speech and Dialogue - TSD'99, Lectures Notes in Artificial Intelligence 1692, Springer-Verlag, 1999, pp. 203-208.

[12] Dubnowski, J.J. & Rabiner *"Real time digital hardware Pitch detector"*. IEEE Transactions on Acoustics speech and signal processing. 24(1), 2-8. 1976

[13] V. Vapnik. The Nature of Statistical Learning Theory, Springer Verlag, New York, 1995.

[14] C. J. C. Burges. *"A Tutorial on Support Vector Machines for Pattern Recognition"* Knowledge Discovery and Data Mining, 2(2), 1998.

15] P.H.Chen, C.J. Lin, B.S.ch¨olkopf*, "A Tutorial on v-Support Vector Machines"* Applied Stochastic Models in Business and Industry 21(2), pages 111-136, 2005.

[16] L. Rabiner and B.H. Juang. Fundamentals of Speech Recognition. Prentice Hall, 1993.