***Yadigar N. Imamverdiyev*[1], *Lyudmila V. Sukhostat*[2]**
Institute of Information Technology of ANAS, Baku, Azerbaijan
[1]yadigar@lan.ab.az, [2]lsuhostat@hotmail.com

DOI: 10.25045/jpit.v07.i2.04

DIALECTS RECOGNITION BASED ON ACOUSTIC MODEL

Regional dialect recognition is important in the field of speech technologies. It is widely used in telephone reference systems, adapting the output synthesized speech in dialog systems, and also in forensics for profiling speaker in judicial or military situations, etc. The article describes different approaches that allow the usage of multiple information sources from the acoustic signal for the construction of dialects recognition system. In particular, acoustic, prosodic, phonetic, and phonotactic approaches are considered.

***Keywords:** dialects recognition, acoustic model, speech signal, UBM-GMM model.*

Introduction

One of the main problems of modern research on speech production and speech technologies is the understanding and modeling of individual variations in the spoken language. Individuals have their own style of speech depending on many factors, such as the dialect and accent, as well as their socio-economic status. These differences tend to exhibit difficulties of speaker-scale simulation systems designed for data processing in a particular language. People have been trying, to a certain extent, to identify and interpret most of these aspects for many years.

Over the past few decades, considerable progress has been observed in automated language identification of the speaker in this speech pattern. The recognition of accent and dialect has only recently begun to attract the attention of scientists in the field of speech technology [1-4]. The problem of dialect identification is the recognition of the regional dialect of the speaker within a given language based on the speech signal.

Modern speech recognition systems focus on speech changes depending on the accent or dialect of a particular language. The dialect of the given language is a model of pronunciation or vocabulary used in the community of native speakers belonging to the same geographical area. Studies show that the use of the speaker dialect recognition before the automated speech recognition improves the system performance by adapting to the respective models (acoustic and language) [5]. This is due to the fact that the speakers who speak different dialects pronounce some words differently, consistently changing certain backgrounds and even morpheme.

Available resources [6, 7] depict that the research on automated dialect identification are performed for different languages of western and eastern countries.

Dialects, accents and styles

There are three different language variations that appear in any language. Two of these categories are determined by regional variations of pronunciation (accent), choice of words, grammatical (dialects) and sociological variations, and in different styles of speech, depending on the age, sex and situation. Knowing all of these variables, a picture of the social, historical and geographical factors of the used language can be imagined [8].

Dialect is a type of speech within the specified language. The differences between dialects mainly occur due to the regional and social factors, and these differences vary in terms of pronunciation, vocabulary and grammar. [7]

Accents are defined as the types of pronunciation of a specific language and refer to the sounds that exist in the language [1]. Unlike dialects, accents cover only a small group of variations that can occur in a particular language.

Styles generally refer to the mood of the speaker and the situation which the speaker is in. This factor differs from the dialect and accent, thus, dialect and accent are the manner of speaking in a particular language in the society, while the styles refer to the spoken language of the same person in different situations.

Dialectology

Dialectology is the study of sounds, words and grammatical forms, varying in language. This term is commonly used to describe the study of accents (differences in the pronunciation of the sounds used in the language) and the dialects (differences in the grammatical structures and words). In general, dialectology focuses on the geographical distribution of different accents and dialects, although it also explores the social factors (such as age, gender, and social status).

A traditional research in dialectology, as a rule, aimed at creating dialect maps, resulting in imaginary lines drawn on the map to indicate the different areas of the dialects. This has led to the study of social and regional language differences. German scientist Johann Andreas Schmeller first explored the Bavarian dialect in 1821-1837, who included linguistic atlas [9]. Linguistic Atlas of the United States (1930s) was one of the first studies of dialects based on the social factors. In 1950, the University of Leeds reviewed English dialects in England and the eastern Wales.

Russian dialects were first studied in the XIX century. "Explanatory Dictionary of Russian language" (1863–1866) by V.I.Dal played a vital role in the study of dialects. First dialectological map of the Russian language was drawn up in 1915 [10, 11].

In Azerbaijan, there are numerous works on the study of the Azerbaijani dialects conducted for many years [12–16].

Dialect Recognition

Various methods have been proposed to solve the problems of the dialect recognition. Most of them are similar to the relevant methods used in language recognition. Furthermore, many studies have been conducted in the field of automated recognition of the regional accent [2, 3, 17–20].

Dialect recognition can be performed on various levels, such as acoustic (e.g., spectral data), prosodic, phonotactical (e.g., language models) and lexical [21]. With regards to acoustic level, the spectral information of the speech signal is extracted through speech parameterization methods, and classification algorithms are then applied, such as a Gaussian mixture model [6], support vector model [22], and neural networks [23]. In [24], an acoustic approach is presented to recognize the four languages of India: Indian, English, Hindi, Assamese, Bengali. Mel-frequency Cepstral Coefficients (MFCCs) are used as parameters, and detection is performed using a Gaussian Mixture Model (GMM). The duration of phonetic units [6, 21, 24, 25] and the rhythm [26] are considered at prosodic level.

In [27, 28], the prosodic parameters are used. Thus, speech signal is divided into unvoiced / voiced segments, wherein the pitch frequency, duration of unvoiced / voiced segments and others is determined. This approach is based on the use of N-gram models. The method aims at separate modeling of phrasal and local accents.

In [29], language recognition system, based on phonotactical approach, is studied. Phonotactical approach in accent and dialect recognition is based on the hypothesis that dialects or accents vary by backgrounds allocation sequences. In other words, the texts in the same language can be recognized by this symbol distribution.

Proposed approach

This paper proposes a GMM-UBM-model to recognize dialects. Universal Background Model (UBM) - is GMM-representation of the characteristics of all the dialects created using part of the training data.

GMM is a density distribution model, which is a language or dialect model. It defines different Gaussian distributions, which have own mathematical expectation, variance and weight in the GMM. Let’s assume that M is a number of small Gaussian distribution models. The following equation is trying to model the probability distribution density N -dimensional random vector *x*, adding a weighted combination of weighted multidimensional Gaussian densities:

$$p(x \mid \lambda_d) = \sum_{i=1}^{M} p_i b_i(x), \tag{1}$$

Where

$$b_i(x) = \frac{1}{(2\pi)^{N/2} \left|\Sigma_i\right|^{1/2}} \exp\left\{\frac{1}{2}(x-\mu_i)' \Sigma_i^{-1} (x-\mu_i)\right\}.$$

μ_i - Vector of mathematical expectations and Σ_i - covariance matrix.

These options are represented as follows:

$$\lambda = \{p_i, \mu_i, \Sigma_i\}, \quad i = 1,...,M \tag{2}$$

where p_i - mixed weight providing condition $\sum_{i=1}^{M} p_i = 1$.

Each group has its own dialects model λ_d.

Model parameters of each dialect are estimated with the use of Expectation-Maximization algorithm (EM) for the sequence of vectors feature $X = \{x_1, x_2,..., x_T\}$, which are derived from a set of speech utterances of the given dialect d.

In addition, for each group of dialects, the following values are found:

$$p_i = \frac{1}{T} \sum_{t=1}^{T} pr(i \left| x_t, \lambda\right.) \tag{3}$$

$$\mu_i = \frac{\sum_{t=1}^{T} pr(i \left| x_t, \lambda\right.) x_t}{\sum_{t=1}^{T} pr(i \left| x_t, \lambda\right.)} , \tag{4}$$

$$\Sigma_i = \frac{\sum_{t=1}^{T} pr(i \left| x_t, \lambda\right.) x_i^2}{\sum_{t=1}^{T} pr(i \left| x_t, \lambda\right.)} - \mu_i^2 , \tag{5}$$

where the posterior probability of the component i has the following form

$$pr(i \left| x_t, \lambda\right.) = \frac{p_i b_i(x)}{\sum_{k=1}^{M} p_k b_k(x)} . \tag{6}$$

GMM-parameters are determined by estimating the maximum probability learning:

$$\lambda_d = \arg\max_{\lambda_i} \{\prod_{t=1}^{T} p(x_t \mid \lambda_i)\} . \tag{7}$$

The advantage of using UBM in dialects identification systems is a significant reduction in the number of training data. Instead of separate education dialect-dependent models, the approach uses Bayesian adaptation of the UBM based on speech samples, trained in a particular dialect. MFCC-coefficients are proposed to be used as parameters. The general scheme of the proposed dialect recognition model is shown in Figure. 1.

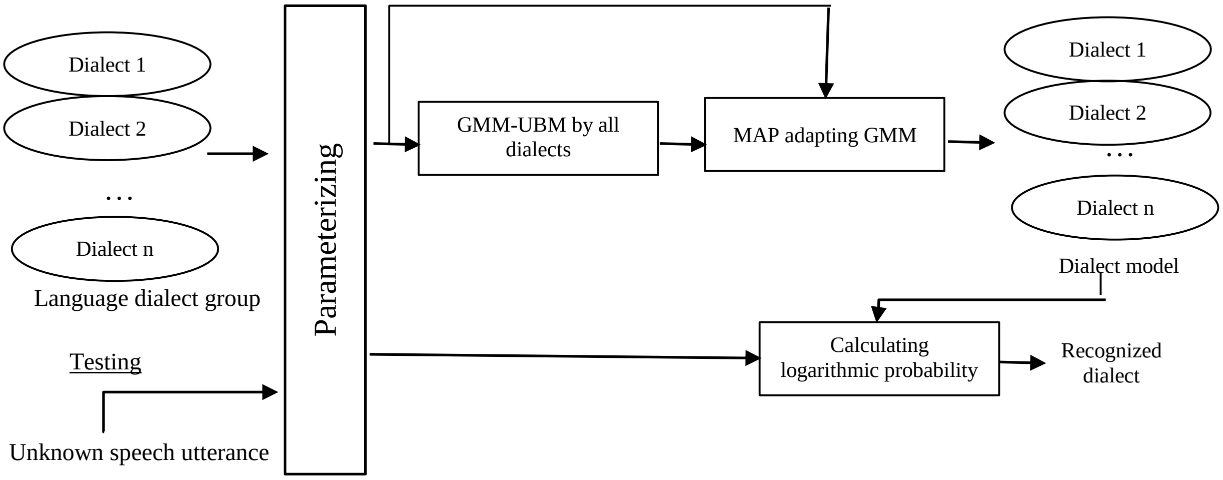

Figure 1. GMM-UBM based dialect identification system

During the identification unknown speech utterance X is classified after calculating the logarithmic probability, which is defined by the formula:

$$L_d(X) = \log p(X \mid \lambda_d) = \frac{1}{T}\sum_{t=1}^{T} \log p(x_t \mid \lambda_d) \ . \qquad (8)$$

Before the automated identification of dialect the speech signals are first pre-processed by filtering at Zero Frequency Filtering (ZFF) [30] to remove low-frequency noise in changing time.

Taking into account that the GMM and UBM have M mixture, we choose the first N mixtures for dialects D to test. For a GMM system, the number of mixed tests is *Nmixture = M × D.*

For the five groups of dialects, using 512 GMM-mixtures the number of tests, is *Nmixture = 512 × 5 = 2560.*

Conclusion

The study is designed to automatically identify the dialects based on speech samples. Acoustic, prosodic, phonetic, and lexical phonotactical approaches are studied to achieve the dialects identification. The model based on GMM is most widely used one. In this article, an acoustic model based on GMM-UBM is proposed to ensure the reliability of recognition.

Thus, we can conclude that the most promising direction of automated dialect recognition is to develop approaches aimed at the initial separation of dialects into the groups followed by the definition of a particular dialect.

References

1. Clark J. An Introduction to phonetics and phonology. Oxford: Blackwell Publishing, 2007, 504 p.
2. Biadsy F. Automatic dialect and accent recognition and its application to speech recognition: Ph.D. dissertation. Columbia University, 2011, 171 p.
3. Omar M.K., Pelecanos J. A novel approach to detecting non-native speakers and their native language / Proc. of IEEE ICASSP, 2010, pp.4398–4401.
4. Roy P., Das P.K. Language ıdentification of Indian languages based on gaussian mixture models // International Journal of Wisdom Based Computing, 2011, vol.1, no.3, pp.54–59.
5. Liu M.K., Xu B., Huang T.Y., Deng Y.G., Li C.R. Mandarin accent adaptation based on context-ındependent/context-dependent pronunciation modeling / Proc. of ICASSP, 2000, vol.2, pp.1025–1028.
6. Hazen, T., Zue, V. Segment-based automatic language ıdentification // Journal of the Acoustic Society of America, 1997, vol.101, no.4, pp.2323–2331.
7. Akmajian A., Demers R.A., Farmer A.K, Harnish R.M. Linguistics: an introduction to

language and communication. Massachusetts: MIT Press, 2001, 604 p.

8. Nerbonne J. Linguistic variation and computation / Proc. of the 10th conference on European chapter of the Association for Computational Linguistics, 2003, vol.1, pp.1–3.
9. Petyt K.M. The study of dialect: an introduction to dialectology. London: Westview Press, 1980, 235 p.
10. Dictionary of Russian folk dialects, vol. 1-43. M. A .: Science, 1965, p. 306.
11. Bromley S.V., Bulatova L.N., Zakharova K.F. Russian dialectology. M .: Education, 1989, p. 223
12. Order of the President of the Republic of Azerbaijan on the "Approval of the State Program on the use of the Azerbaijani language in accordance with the requirements of globalization and the development of linguists in the country", "Azerbaijan" newspaper, Baku, 2013, April 9.
13. Shiraliyev M.Sh. Dialectic fundamentals of the Azerbaijani national literal language. M .: Publishing House of the USSR Academy of Sciences, 1958, pp.78–84.
14. Dialectological dictionary of the Azerbaijani language. Baku: "East-West", 2007, p.568
15. Bayramov I. Vocabulary of the dialect of the Western Azerbaijan. Baku: "Science and Education", 2011, p.440
16. Kazimov G. The study of the history of dialects and accents // Problems of philology, 2014, no.1, pp.3–27.
17. Teixeira C., Trancoso I., Serralheiro A. Accent identification / Proc. of INTERSPEECH, 1996, vol.3, pp.1784–1787.
18. Hanani A., Russell M.J., Carey M.J. Human and computer recognition of regional accents and ethnic groups from british english speech //Computer Speech and Language, 2013, vol.27, no.1, pp.59–74.
19. Huang R., Hansen J.H.L., Angkititrakul P. Dialect/accent classification using unrestricted audio // IEEE Trans. on Audio, Speech and Language Processing, 2007, vol.15, no.2, pp.453–464.
20. Mporas I., Ganchev T., Fakotakis N. Phonotactic recognition of greek and cypriot dialects from telephone speech // SETN 2008, Advances in Artificial Intelligence, Lecture Notes in Computer Science. Berlin: Springer-Verlag, 2008, pp.173–181.
21. Tong, R., Ma, B., Li, H., Chng, E.S. Integrating acoustic, prosodic and phonotactic features for spoken language ıdentification / Proc. of IEEE ICASSP, 2006, pp.205–208.
22. Campbell W. M., Singer E., Torres-Carrasquillo P.A., Reynolds D.A. Language recognition with support vector machines / Proc. of Odyssey, 2004, pp.285–288.
23. Braun J., Levkowitz H. Automatic language ıdentification with perceptually guided training and recurrent neural networks / Proc. of the 5th International Conference on Spoken Language Processing (ICSLP), 1998, pp.3201–3205.
24. Ghesquiere P.J., Compernolle D.V. Flemish accent ıdentification based on formant and duration features // Proc. of ICASSP, 2002, pp.749–752.
25. Lin C.Y., Wang H.C. Fusion of phonotactic and prosodic knowledge for language ıdentification / Proc. of the 9th International Conference on Spoken Language Processing (ICSLP), 2006, pp.425–428.
26. Farinas J., Pellegrino F., Rouas J.L., Andre-Obrecht R. Merging segmental and rhythm features for automatic language identification / Proc. of IEEE ICASSP, 2002, pp.753–756.
27. Rouas J.-L., Farinas J., Pellegrino F., Andre-Obrecht R. Modeling prosody for language identification on read and spontaneous speech / IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003, vol.6, pp.1–4.
28. Rouas J-L. Automatic prosodic variations modeling for language and dialect discrimination // IEEE transactions on audio, speech and language processing, 2007, vol.15, no.6, pp.1904–1911.
29. Torres-Carrasquillo P. A., Singer E., Kohler M. A. Approaches to language ıdentification using gaussian mixture models and shifted delta cepstral features / Proc. of IEEE ICASSP, 2002, pp.757–760.
30. Murty K. S. R., Yegnanarayana B. Epoch extraction from speech signals // IEEE Transactions on Audio, Speech, and Language Processing, 2008, vol.16, no.8, pp.1602–1613.