ISSN 0735-2727, Radioelectronics and Communications Systems, 2020, Vol. 63, No. 1, pp. 42–54. © Allerton Press, Inc., 2020.
Russian Text © The Author(s), 2020, published in Izvestiya Vysshikh Uchebnykh Zavedenii, Radioelektronika, 2020, Vol. 63, No. 1, pp. 55–68.

Minimum of Information Divergence Criterion for Signals with Tuning to Speaker Voice in Automatic Speech Recognition

V. V. Savchenko[1]

[1]*Nizhny Novgorod State Linguistic University, Nizhny Novgorod, Russia*
ORCID: 0000-0003-3045-3337, e-mail: vvsavchenko@yandex.ru

Received May 16, 2019
Revised September 12, 2019
Accepted December 12, 2019

Abstract—It is considered a problem of automatic speech recognition at basic, phonetic level of speech signal processing. It is researched a problem of noise-immunity increase. For its solution it is proposed a criterion of minimum information divergence of the signals with tuning to a speaker voice and automatic scaling of speech template to thin structure of observed (current) speech frame. An example of its practical realization is considered, efficiency characteristics are researched. Applying the authors' software we carry out an experiment and obtain qualitative estimations of threshold signals gain in case of application of proposed criterion. It is shown than this gain can be 10 dB and greater under certain conditions. Obtained results and drawn conclusions are intended it to their application for development and modernization of existent systems and techniques of automatic processing and recognition of speech intended it to operation in conditions of intensive noise effect.

DOI: 10.3103/S0735272720010045

INTRODUCTION

Automated speech recognition (ASR) is one of the most developing directions in digital processing of signals for many years [1]. The most popular methods are neural networks methods and models [2], on their basis using technology "client-server" there are realized the most known developed methods [2, 3]. In this trending direction of ASR now there are achieved the most impressive results including commercial.

But there are amount of problems which put obstacles in the way of further progress in this research field. General and principal drawback of multi-layer neural network structures is their resource-intensity generating a problem [4] of practical realization of current-technology ASR algorithms in autonomous (without internet application) version. There is related another actual problem in conditions of information society: protection of voice information from unauthorized access [5]. There is a daunting problem especially in the systems of voice control [6]. Hence, last years ASR practical application field is approximately in the limits of interactive inquiry-information systems [3] that does not meet the potential possibilities of voice technologies.

Efficient way of voice information protection is tuning of the ASR algorithm to the user voice [1, 5]. But such tuning (not confuse the system adaptation to specificities of the user voice [2]) is principally inapplicable in case of neural networks approach to ASR, departing from the principle of independence of decision-making form the speakers.

Therefore many specialists use [5, 6] recourse-saving phonetic approach [1, 7], which is based on probabilistic models of minimum speech units (MSU) like phonemes and their allophones. At that its application field does not restricted by voice control systems and it spreads to coding systems and voice transmission via leased channel and also analysis of thin structure of voice signals in the problems of voice identification of speakers, users verification, language learning, etc. [8, 9].

ASR problem in general case is stated [9, 10] as multi-alternative verification of statistic hypotheses about a law of probabilities distribution of samples vector of observed (current) MSU. Tuning to a speaker voice is realized by means of shaping of finite set of reference distributions at preparatory stage. Mentioned procedure is not a problem in calculation meaning [8] and it can be used in the mode of regular update of phonetic data base form specific speaker calculating on known variations of phonetic system of his oral

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ADDITIONAL INFORMATION

The initial version of this paper in Russian is published in the journal "Izvestiya Vysshikh Uchebnykh Zavedenii. Radioelektronika," ISSN 2307-6011 (Online), ISSN 0021-3470 (Print) on the link http://radio.kpi.ua/article/view/S0021347020010045 with DOI: 10.20535/S0021347020010045.

REFERENCES

1. L. R. Rabiner, R. W. Shafer, *Theory and Applications of Digital Speech Processing* (Pearson, Boston, 2010). URI: https://www.pearson.com/us/higher-education/program/Rabiner-Theory-and-Applications-of-Digital-Speech-Processing/PGM130812.html.
2. I. B. Tampel, "Automated speech recognition – the main stages over last 50 years," *Sci. Tech. J. Information Technol., Mech. Optics* **15**, No. 6, 957 (2015). DOI: 10.17586/2226-1494-2015-15-6-957-968.
3. M. Schuster, "Speech recognition for mobile devices at Google," in: B. T. Zhang, M. A. Orgun (eds.), *PRICAI 2010: Trends in Artificial Intelligence*. PRICAI 2010. Lecture Notes in Computer Science (Springer, Berlin, Heidelberg, 2010), Vol. 6230, pp. 8-10. DOI: 10.1007/978-3-642-15246-7_3.
4. V. V. Savchenko, A. V. Savchenko, "Information-theoretic analysis of efficiency of the phonetic encoding-decoding method in automatic speech recognition," *J. Commun. Technol. Electronics* **61**, No. 4, 430 (2016). DOI: 10.1134/S1064226916040112.
5. Z. Wu, *Information Hiding in Speech Signals for Secure Communication* (Elsevier Science, 2015). DOI: 10.1016/C2013-0-19179-9.
6. R. Rammohan, N. Dhanabalsamy, V. Dimov, J. Frank, "Eidelman smartphone conversational agents (Apple Siri, Google, Windows Cortana) and questions about allergy and asthma emergencies," *J. Allergy Clinical Immunology* **139**, No. 2, ab250 (2017). DOI: 10.1016/j.jaci.2016.12.804.
7. M. B. Akçay, K. Oğuzb, "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities and classifiers," *Speech Commun.* **116**, No. 1, 56 (2020). DOI: 10.1016/j.specom.2019.12.001.
8. V. V. Savchenko, "A method of measuring the index of acoustic voice quality based on an information-theoretic approach," *Meas. Tech*. **61**, No. 1, 79 (2018). DOI: 10.1007/s11018-018-1391-8.
9. V. V. Savchenko, "Itakura-Saito divergence as an element of the information theory of speech perception," *J. Commun. Technol. Electron.* **64**, No. 6, 590 (2019). DOI: 10.1134/S1064226919060093.
10. V. V. Savchenko, "Criterion for minimum of mean information deviation for distinguishing random signals with similar characteristics," *Radioelectron. Commun. Syst.* **61**, No. 9, 419 (2018). DOI: 10.3103/S0735272718090042.
11. S. M. Qaisar, N. Hammad, R. Khan, R. Asfour, "A speech to machine interface based on perceptual linear prediction and classification," *Proc. of Int. Conf. on Advances in Science and Engineering Technology*, 26 Mar.-10 Apr. 2019, Dubai, UAE (IEEE, 2019). DOI: 10.1109/ICASET.2019.8714304.
12. V. N. Zvaritch, B. G. Marchenko, "Linear autoregressive processes with periodic structures as models of information signals," *Radioelectron. Commun. Syst.* **54**, No. 7, 367 (2011). DOI: 10.3103/S0735272711070041.
13. F. Castanié, *Digital Spectral Analysis: Parametric*, *Non-Parametric and Advanced Methods* (Wiley-ISTE, 2011). DOI: 10.1002/9781118601877.
14. V. V. Savchenko, A. V. Savchenko, "Criterion of significance level for selection of order of spectral estimation of entropy maximum," *Radioelectron. Commun. Syst.* **62**, No. 5, 223 (2019). DOI: 10.3103/S0735272719050042.
15. R. M. Gray, A. Buzo, A. H. Gray, Y. Matsuyama, "Distortion measures for speech processing," *IEEE Trans. Acoust.*, *Speech Signal Processing* **28**, No. 4, 367 (1980). DOI: 10.1109/TASSP.1980.1163421.
16. O. D. Eva, A. M. Lazar, "Feature extraction and classification methods for a motor task brain computer interface: a comparative evaluation for two databases," *Int. J. Advanced Computer Sci. Appl*. **8**, No. 8, 263 (2017). DOI: 10.14569/IJACSA.2017.080834.
17. S. S. Rachel, U. Snekhalatha, K. Vedhasorubini, D. Balakrishnan, "Spectral analysis of speech signal characteristics: a comparison between healthy controls and laryngeal disorder," *Proc. of Int. Conf. on Intelligent Computing and Applications* (Springer, Singapore, 2018), Vol. 632, pp. 333-334. DOI: 10.1007/978-981-10-5520-1_31.
18. V. V. Savchenko, "Words phonetic decoding method with the suppression of background noise," *J. Commun. Technol. Electron*. **62**, No. 7, 788 (2017). DOI: 10.1134/S1064226917070099.
19. E. Hossain, M. S. A. Zilany, E. Davies-Venn, "On the feasibility of using a bispectral measure as a nonintrusive predictor of speech intelligibility," *Computer Speech Lang.* **57**, 59 (2019). DOI: 10.1016/j.csl.2019.02.003.

20. H. Ding, T. Lee, I. Y. Soon, C. K. Yeo, P. Dai, G. Dan, “Objective measures for quality assessment of noise-suppressed speech,” *Speech Commun.* **71**, 62 (2015). DOI: 10.1016/j.specom.2015.02.001.
21. A. A. Borovkov, *Mathematic Statistics* [in Russian] (Lan’, St. Petersburg, 2010).
22. S. Kullback, *Information Theory and Statistics* (Dover Pub., N.Y., 1997).
23. E. Estrada, H. Nazeran, F. Ebrahimi, M. Mikaeili, “Symmetric Itakura distance as an EEG signal feature for sleep depth determination,” *Proc. of ASME Bioengineering Conf.*, 17-21 Jun. 2009, Lake Tahoe, USA (2009), pp. 723-724. DOI: 10.1115/SBC2009-206233.
24. A. A. Gharbali, S. Najdi, J. M. Fonseca, “Investigating the contribution of distance-based features to automatic sleep stage classification,” *Comput. Biology Medicine* **96**, 8 (2017). DOI: 10.1016/j.comp-biomed.2018.03.001.
25. B. R. Levin, *Theoretic Principles of Statisitc Radioengineering* [in Russian] (Radio i Svyaz’, Moscow, 1989).