

УДК 621.391:004.934

КРИТЕРИЙ МИНИМУМА ИНФОРМАЦИОННОГО РАСХОЖДЕНИЯ СИГНАЛОВ С НАСТРОЙКОЙ НА ГОЛОС ДИКТОРА В ЗАДАЧЕ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ

В. В. САВЧЕНКО

*Нижегородский государственный лингвистический университет,
Россия, Нижний Новгород, 603155, ул. Минина, 31а*

Аннотация. Рассмотрена задача автоматического распознавания речи на базовом, фонетическом уровне обработки речевого сигнала. Исследована проблема повышения помехоустойчивости. Для ее решения предложен критерий минимума информационного расхождения сигналов с настройкой на голос диктора и с автоматическим масштабированием речевых эталонов под тонкую структуру наблюдаемого (текущего) речевого фрейма. Рассмотрен пример его практической реализации, исследованы характеристики эффективности. С использованием авторского программного обеспечения поставлен и проведен эксперимент, получены количественные оценки выигрыша в пороговых сигналах. Показано, что при определенных условиях он может достигать 10 дБ и более. Полученные результаты и сделанные по ним выводы предназначены для использования при разработке новых и модернизации существующих систем и технологий автоматической обработки и распознавания речи, рассчитанных на работу в условиях действия интенсивных внешних помех.

Ключевые слова: случайный сигнал; цифровая обработка сигналов; речевой сигнал; автоматическая обработка речи; речевые технологии; помехоустойчивость

ВВЕДЕНИЕ

На протяжении многих лет автоматическое распознавание речи (АРР) относится к числу наиболее динамично развивающихся направлений в области цифровой обработки сигналов [1]. В методологическом отношении здесь доминируют нейросетевые методы и модели [2], на базе которых по технологии «клиент-сервер» выполнены все наиболее известные в мире разработки [2, 3]. Именно в этом трендовом направлении АРР в настоящее время достигнуты наиболее впечатляющие, в том числе с коммерческой точки зрения, результаты.

Однако существует ряд проблем, служащих препятствием для дальнейшего прогресса в данной области исследований. Общий и

принципиальный недостаток многослойных нейросетевых структур — их многоресурсность, порождающая собой проблему [4] практической реализуемости современных алгоритмов АРР в автономном (без выхода в интернет) варианте. С ней тесно связана еще одна актуальная в условиях информационного общества проблема: защиты речевой информации от несанкционированного доступа [5]. Особенно остро она проявляется в системах голосового управления [6]. Как следствие, область практического применения АРР в течение многих лет почти не выходит за рамки интерактивных справочно-информационных систем [3], что, безусловно, не отвечает потенциальным возможностям речевых технологий.

DOI: [10.20535/S0021347020010045](https://doi.org/10.20535/S0021347020010045)
© В. В. Савченко, 2020

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Rabiner, L. R.; Shafer, R. W. *Theory and Applications of Digital Speech Processing*. Boston: Pearson, 2010. URI: <https://www.pearson.com/us/higher-education/program/Rabiner-Theory-and-Applications-of-Digital-Speech-Processing/PGM130812.html>.
2. Тампель, И. Б. "Автоматическое распознавание речи — основные этапы за 50 лет," *Научно-технический вестник информационных технологий, механики и оптики*, Т. 15, № 6, С. 957-968, 2015. DOI: [10.17586/2226-1494-2015-15-6-957-968](https://doi.org/10.17586/2226-1494-2015-15-6-957-968).
3. Schuster, M. "Speech recognition for mobile devices at Google," in: Zhang, B T.; Orgun, M. A. (eds.) *PRICAI 2010: Trends in Artificial Intelligence*. PRICAI 2010. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, Vol. 6230. p. 8-10, 2010. DOI: [10.1007/978-3-642-15246-7_3](https://doi.org/10.1007/978-3-642-15246-7_3).
4. Savchenko, V. V.; Savchenko, A. V. "Information-theoretic analysis of efficiency of the phonetic encoding-decoding method in automatic speech recognition," *J. Commun. Technol. Electronics*, Vol. 61, No. 4, p. 430-435, 2016. DOI: [10.1134/S1064226916040112](https://doi.org/10.1134/S1064226916040112).
5. Wu, Z. *Information Hiding in Speech Signals for Secure Communication*. Elsevier Science, 2015. DOI: [10.1016/C2013-0-19179-9](https://doi.org/10.1016/C2013-0-19179-9).
6. Rammohan, R.; Dhanabalsamy, N.; Dimov, V.; Frank J. "Eidelman smartphone conversational agents (Apple Siri, Google, Windows Cortana) and questions about allergy and asthma emergencies," *Journal of Allergy and Clinical Immunology*, Vol. 139, Iss. 2, p. ab250, 2017. DOI: [10.1016/j.jaci.2016.12.804](https://doi.org/10.1016/j.jaci.2016.12.804).
7. Akçay, M. B.; Oğuzb, K. "Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities and classifiers," *Speech Communication*, Vol. 116, No.1, p. 56-76, 2020. DOI: [10.1016/j.specom.2019.12.001](https://doi.org/10.1016/j.specom.2019.12.001).
8. Savchenko, V. V. "A method of measuring the index of acoustic voice quality based on an information-theoretic approach," *Meas. Tech.*, Vol. 61, No. 1, p. 79-84, 2018. DOI: [10.1007/s11018-018-1391-8](https://doi.org/10.1007/s11018-018-1391-8).
9. Savchenko, V. V. "Itakura-Saito divergence as an element of the information theory of speech perception," *J. Commun. Technol. Electron.*, Vol. 64, No. 6, p. 590-596, 2019. DOI: [10.1134/S1064226919060093](https://doi.org/10.1134/S1064226919060093).
10. Савченко, В. В. "Критерий минимума среднего информационного отклонения для различения случайных сигналов с близкими характеристиками," *Известия вузов. Радиоэлектроника*, Т. 61, № 9, С. 536-547, 2018. DOI: [10.20535/S0021347018090042](https://doi.org/10.20535/S0021347018090042).
11. Qaisar, S. M.; Hammad, N.; Khan, R.; Asfour, R. "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](https://doi.org/10.1109/ICASET.2019.8714304).

12. Зварич, В.Н.; Марченко, Б.Г. "Линейные процессы авторегрессии с периодическими структурами как модели информационных сигналов," *Известия вузов. Радиоэлектроника*, Т. 54, № 7, С. 25-30, 2011. DOI: [10.20535/S0021347011070041](https://doi.org/10.20535/S0021347011070041).

13. Castanié, F. *Digital Spectral Analysis: Parametric, Non-Parametric and Advanced Methods*. Wiley-ISTE, 2011. DOI: [10.1002/9781118601877](https://doi.org/10.1002/9781118601877).

14. Савченко, В.В.; Савченко А.В. "Критерий регулируемого уровня значимости для выбора порядка спектральной оценки максимума энтропии," *Известия вузов. Радиоэлектроника*, Т. 62, № 5, С. 276-286, 2019. DOI: [10.20535/S0021347019050042](https://doi.org/10.20535/S0021347019050042).

15. Gray, R. M.; Buzo, A.; Gray, A. H.; Matsuyama, Y. "Distortion measures for speech processing," *IEEE Trans. Acoust., Speech Signal Processing*, Vol. 28, No. 4, p. 367-376, 1980. DOI: [10.1109/TASSP.1980.1163421](https://doi.org/10.1109/TASSP.1980.1163421).

16. Eva, O. D.; Lazar, A. M. "Feature extraction and classification methods for a motor task brain computer interface: a comparative evaluation for two databases," *Int. J. Advanced Computer Sci. Appl.*, Vol. 8, No. 8, p. 263-269, 2017. DOI: [10.14569/IJACSA.2017.080834](https://doi.org/10.14569/IJACSA.2017.080834).

17. Rachel, S. S.; Snehalatha, U.; Vedhasorubini, K.; Balakrishnan, D. "Spectral analysis of speech signal characteristics: a comparison between healthy controls and laryngeal disorder," *Proc. of Int. Conf. on Intelligent Computing and Applications*. Singapore: Springer, Vol. 632, p. 333-334, 2018. DOI: [10.1007/978-981-10-5520-1_31](https://doi.org/10.1007/978-981-10-5520-1_31).

18. Savchenko, V. V. "Words phonetic decoding method with the suppression of background noise," *J. Commun. Technol. Electron.*, Vol. 62, No. 7, p. 788-793, 2017. DOI: [10.1134/S1064226917070099](https://doi.org/10.1134/S1064226917070099).

19. Hossain, E.; Zilany M.S.A.; Davies-Venn, E. "On the feasibility of using a bispectral measure as a nonintrusive predictor of speech intelligibility," *Computer Speech Lang.*, Vol. 57, p. 59-80, 2019. DOI: [10.1016/j.csl.2019.02.003](https://doi.org/10.1016/j.csl.2019.02.003).

20. Ding, H.; Lee, T.; Soon I. Y.; Yeo, C. K.; Dai, P.; Dan, G. "Objective measures for quality assessment of noise-suppressed speech," *Speech Commun.*, Vol. 71, p. 62-73, 2015. DOI: [10.1016/j.specom.2015.02.001](https://doi.org/10.1016/j.specom.2015.02.001).

21. Боровков, А. А. *Математическая статистика*. СПб.: Лань, 2010. 704 с.

22. Кульбак, С. *Теория информации и статистика*: Пер. с англ. М.: Наука, 1967. 408 с.

23. Estrada, E.; Nazeran, H.; Ebrahimi, F.; Mikaeili, M. "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, p. 723-724. DOI: [10.1115/SBC2009-206233](https://doi.org/10.1115/SBC2009-206233).

24. Gharbali, A. A.; Najdi, S.; Fonseca, J. M. "Investigating the contribution of distance-based features to automatic sleep stage classification," *Comput. Biology Medicine*, Vol. 96, p. 8-23, 2017. DOI: [10.1016/j.compbiomed.2018.03.001](https://doi.org/10.1016/j.compbiomed.2018.03.001).

25. Левин, Б.Р. *Теоретические основы статистической радиотехники*, изд. 3-е, перераб. и доп. М.: Радио и связь, 1989. 656 с.

Поступила в редакцию 16.05.2019

После доработки 21.10.2019

Принята к публикации 25.12.2019