

АНАЛИТИЧЕСКИЙ ОБЗОР И КЛАССИФИКАЦИЯ МЕТОДОВ ВЫДЕЛЕНИЯ ПРИЗНАКОВ АКУСТИЧЕСКОГО СИГНАЛА В РЕЧЕВЫХ СИСТЕМАХ

И.А. ГУРТУЕВА, К.Ч. БЖИХАТЛОВ

Институт информатики и проблем регионального управления –
филиал Кабардино-Балкарского научного центра Российской академии наук
360000, Россия, Нальчик, ул. И. Арманд, 37-а

Аннотация. В данной работе представлен обзор методов и алгоритмов выделения признаков при трансформации акустического сигнала в последовательность векторов для решения задач сегментации, классификации, идентификации или распознавания речи. Предложена классификация методов извлечения признаков по математическим подходам. Обсуждаются алгоритмы и техники спектрального анализа, наиболее применяемые при проектировании систем распознавания речи. Настоящий обзор наглядно демонстрирует сложность проблемы акустической обработки – отыскания представления, снижающего размерность модели при сохранении полноты лингвистической информации и, что важно, устойчивого к вариативности относительно диктора, каналов передачи и окружающей среды. Проведенный анализ существующих методов извлечения признаков полезен для выбора технологии при проектировании ключевого элемента речевой системы.

Ключевые слова: распознавание речи, преобразование Фурье, кепстральный анализ, линейное предсказание, методы выделения признаков

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Deller J.R., Proakis J.G., Hansen J. H.L. Discrete Time Processing of Speech Signals. Hoboken NJ: Wiley-IEEE Press, 1999. 936 p.
2. Gupta V. A Survey of Natural Language Processing Techniques. International Journal of Computer Science & Engineering Technology (IJCSET). 2014. Vol. 5. No. 1. Pp. 14–16.
3. Picone J.W. Signal modeling techniques in speech recognition. Proceedings of the IEEE. 1993. Vol. 81. No. 9. Pp. 1215–1245.
4. Jurafsky D., Martin J. Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition. NJ: Prentice Hall, 2009. 1024 p.
5. Pickles J.O. An Introduction to the Physiology of Hearing. New York: Academic Press, 1988. 400 p.
6. Fletcher H., Munson W.A. Relation between Loudness and Masking. J. Acoust. Soc. Am. 1937. No. 9. Pp. 1–10.
7. Zviker E., Feldkeller R. *Ukho kak priyemnik informatsii* [Ear as a receiver of information]. Moscow: Svyaz', 1971. 255 p. (In Russian)
8. Schroeder M.R. Optimizing Digital Speech Coders by Exploiting Masking properties of the Human Ear. J. Acoust. Soc. Am. 1979. No. 66(6). Pp. 1647–1652. <https://doi.org/10.1121/1.383662>
9. Traunmuller H.: Analytical Expressions for the tonotopic sensory scale. The Journal of the Acoustical Society of America. 1990. Vol. 88. No 1. Pp. 97–100. DOI: 10.1121/1.399849.
10. Kavalchuk A.N. The formula for the transition from the frequency domain to the bark scale and vice versa. *Informatika* [Informatics]. 2011. No. 4(32). Pp. 71–81. (In Russian)
11. Aldoshina I., Pritts R. *Muzykal'naya akustika* [Musical acoustics]. Textbook. St. Petersburg: Composer, 2014. P. 720. (In Russian)
12. Moore B. Frequency selectivity in hearing. Boston, MA: Springer, 1986. P. 456. <https://doi.org/10.1007/978-1-4613-2247-4>

13. *Smith J.O.* Mathematics of the Discrete Fourier Transform (DFT) with Audio Applications. W3K Publishing. 2007. P. 322 <http://books.w3k.org/>
14. *Rabiner L.R., Schafer R.W.* Digital processing of speech signal. New Jersey: Prentice-Hall, 1978. P. 496 (Russ. ed.: *Rabiner L.R., Shafer R.V. Tsifrovaya obrabotka rechevykh signalov.* Moscow: Radio i svyaz' Publ., 1981. 496 p.).
15. *Davis S., Mermelstein P.* Experiments in syllable-based recognition of continuous speech. IEEE Transactions on Acoustics, Speech and Signal Processing. 1980. Vol. 28. Pp. 357–366.
16. *Chakraborty S., Roy A., Saha G.* Fusion of a complementary feature set with MFCC for improved closed set text-independent speaker identification. IEEE International Conference on Industrial Technology. Mumbai, 2006. Pp. 387–390.
17. *Shao Y., Wang D.L.* Robust speaker identification using auditory features and computational auditory scene analysis. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008). Las Vegas, NV, USA, 2008. Pp. 1589–1592. DOI: 10.1109/ICASSP.2008.4517928.
18. *Novikov L.V. Osnovy veyvlet-analiza signalov [Fundamentals of Wavelet Signal Analysis].* Tutorial. St. Petersburg: IanP RAN, 1999. 152 p. ([In Russian](#))
19. *Huang N.E.* The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London A. 1998. Vol. 454. Pp. 903–995.
20. *Todisco M., Delgado H., Evans N.* A new feature for automatic speaker verification anti-spoofing: constant Q cepstral coefficients. Odyssey 2016. The Speaker and Language Recognition Workshop. Bilbao, Spain, 2016. Pp. 283–290.
21. *Fant G.* Acoustic Theory of Speech Production. Walter de Gruyter, 1970. P. 328.
22. *Rabiner L., Juang B.-H.* Fundamentals of speech recognition. NJ: Prentice-Hall, Inc., 1993. P. 507.

Информация об авторах

Гуртуева Ирина Асланбековна, науч. сотр. отдела «Компьютерная лингвистика», Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра РАН;

360000, Россия, Нальчик, ул. И. Арманд, 37-а;

gurtueva-i@yandex.ru, ORCID: <https://orcid.org/0000-0003-4945-5682>

Бжихатлов Кантемир Чамалович, канд. физ.-мат. наук, зам. директора по науке, Институт информатики и проблем регионального управления – филиал Кабардино-Балкарского научного центра РАН;

360000, Россия, Нальчик, ул. И. Арманд, 37-а;

haosit13@mail.ru, ORCID: <https://orcid.org/0000-0003-0924-0193>