

UO‘K 004.934.8

## DELTA-GATED SPIKE ENCODING (DGSE) ASOSIDAGI SIYRAK VA ENERGIYA-TEJAMKOR BELGILAR TO‘PLAMINI AJRATISH USULI

<sup>+</sup> *Nurimov P.B.<sup>1</sup>, Babomuradov O.J.<sup>2</sup>*

<sup>1</sup> “TIQXMMI” Milliy tadqiqot universiteti, Toshkent, O‘zbekiston

<sup>2</sup> Qozon federal universiteti Jizzax filiali, Jizzax, O‘zbekiston

<sup>+</sup> [paranur87@gmail.com](mailto:paranur87@gmail.com)

**Annotatsiya.** Mazkur maqolada shaxsni ovozi orqali tanib olish masalasi uchun yangi feature extraction usuli – Delta-Gated Spike Encoding (DGSE) taklif qilinadi. Taklif etilgan yondashuv log-Mel spektrogramma, temporal delta, adaptiv threshold, energy gate va musbat/manfiy spike kodlash bosqichlarini birlashtiradi. Maqsad – akustik signaldagi informativ vaqt-chastota o‘zgarishlarini ajratib olib, siyrak va keyingi spiking yoki energiya-tejamkor modellar uchun qulay ko‘rinishdagi ifodani hosil qilishdan iborat. Tajribalar VoxCeleb1 datasetida uch bosqichli parametrik qidiruv asosida amalga oshirildi. Dastlabki qo‘pol qidiruv, keyingi kengaytirilgan qidiruv va yakuniy fine search natijalari taqqoslandi. Eng yaxshi natija  $\alpha = 1.0$ ,  $\beta = 0.05$ ,  $\text{energy\_thr} = -5.25$  parametrlarida olindi. Ushbu holatda total spike rate = 0.079585, sparsity = 0.920415, gate open rate = 0.718533 ga teng bo‘ldi. Olingan natijalar DGSE usuli signalning informativ qismlarini saqlagan holda yuqori siyraklikni ta‘minlashini ko‘rsatadi. Bu esa uni resursi cheklangan qurilmalar hamda CNN-SNN gibrid shaxsni ovozi orqali tanib olish tizimlari uchun istiqbolli yechimga aylantiradi.

**Kalit so‘zlar:** speaker recognition, feature extraction, log-Mel, spike encoding, sparsity, neuromorfik hisoblash, VoxCeleb1, DGSE.

### 1 KIRISH

Shaxsni ovozi orqali tanib olish biometrik autentifikatsiya, verifikatsiya va identifikatsiya tizimlarida keng qo‘llaniladi; bu yo‘nalishning usullari va rivojlanish bosqichlari bo‘yicha umumiy tahlillar mavjud [3, 10].

Nutq signalidan belgi ajratishda MFCC va log-Mel kabi klassik akustik ifodalar uzoq yillardan beri samarali yechim sifatida ishlatilib kelinmoqda [1, 2]. Mel-chastotali parametrik ifodaning amaliy samaradorligi Davis va Mermelstein ishida ko‘rsatilgan, temporal dinamik belgilar foydasi esa Furui tomonidan yoritilgan.

So‘nggi yillarda x-vector va ECAPA-TDNN kabi chuqur o‘rganishga asoslangan embedding usullari speaker verification sohasida kuchli baseline sifatida qaralmoqda [6, 7]. Bundan tashqari, resurslari cheklangan qurilmalar uchun siyrak va event-driven hisoblashga asoslangan neuromorfik yondashuvlarga qiziqish ortmoqda [9].

Aksariyat zamonaviy shaxsni ovozi orqali tanib olish tizimlari zich akustik representationlar bilan ishlaydi. Bunday ifodalar yuqori aniqlik berishi mumkin, ammo ularda redundant vaqt-chastota ma‘lumotlari ko‘p bo‘ladi. Ayniqsa, edge qurilmalar yoki energiya cheklangan platformalarda bu hisoblash xarajatini oshiradi. Shu sababli faqat informativ o‘zgarishlarni ajratib oladigan, siyrak va hodisaga asoslangan front-end yaratish dolzarb masaladir.

Mazkur ishda Delta-Gated Spike Encoding (DGSE) taklif qilinadi. Ushbu usul log-Mel spektrogramma asosida temporal deltani hisoblaydi, adaptiv threshold qo‘llaydi, energy gate yordamida kam informativ segmentlarni filtrlab, musbat va manfiy spike larni generatsiya qiladi. Yondashuvning asosiy g‘oyasi – butun signalni emas, balki speakerga oid muhim dinamik o‘zgarishlarni event ko‘rinishda ifodalash.

Mazkur ishning asosiy hissasi quyidagilardan iborat:

- shaxsni ovozi orqali tanib olish uchun yangi spike-oriented feature extraction usuli taklif qilinadi;
- temporal delta, adaptiv threshold va energy gate yagona pipeline ichida birlashtiriladi;
- VoxCeleb1 datasetida uch bosqichli parametrik qidiruv asosida optimal parametrlar aniqlanadi;
- representationning siyraklik va spike faolligi bo‘yicha xususiyatlari tahlil qilinadi.

MFCC va Mel asosidagi ifodalar nutqni qayta ishlashda eng ko‘p ishlatiladigan klassik belgilardandir. Davis va Mermelstein tomonidan tavsiya qilingan parametrik yondashuv Mel-chastota shkalasining amaliy qiymatini ko‘rsatib berdi [1].

Temporal o'zgarishlarga asoslangan delta belgilar esa spektral dinamika orqali qo'shimcha axborot berishi mumkinligi sababli keyingi ishlarda keng qo'llanila boshladi. Bu yo'nalish Furui ishlarida ancha erta asoslab berilgan [2, 3].

Shaxsni ovozi orqali tanib olish bo'yicha yirik ochiq datasetlardan biri VoxCeleb bo'lib, u "in the wild" yozuvlarni o'z ichiga oladi va 2017-yildagi ishda yirik speaker identification korpusi sifatida taqdim etilgan. Keyinchalik VoxCeleb2 undan ham kattaroq hajmda taqdim etilib, 1 milliondan ortiq utterance va 6000 dan ortiq speaker ni qamrab olgani qayd etilgan [4, 5].

Chuqur o'rganish asosidagi zamonaviy speaker embedding usullari orasida x-vector va ECAPA-TDNN alohida o'rin tutadi. X-vector DNN embedding yondashuvi sifatida shaxsni ovozi orqali tanib olish uchun mustahkam baseline bo'lib qoldi. ECAPA-TDNN esa TDNN arxitekturasini channel attention, propagation va aggregation bilan kuchaytirib, VoxCeleb benchmarklarida kuchli natijalar ko'rsatgan [6, 7].

Shu bilan birga, neuromorfik hisoblash va spiking neural networks bo'yicha tadqiqotlar event-driven, siyrak va energiya-tejamkor hisoblashni rivojlantirmoqda. Bu shaxsni ovozi orqali tanib olish uchun ham spike-oriented front-end larni ishlab chiqish zaruratini kuchaytiradi [10].

Mavjud ishlardan farqli ravishda, taklif etilayotgan DGSE klassik log-Mel representationni saqlab qolgan holda temporal delta, adaptiv threshold va energy gate ni birlashtiradi va natijani musbat/manfiy spike lar ko'rinishida beradi. Shu jihati bilan u oddiy threshold spike kodlashdan farq qiladi.

## 2 MATERIALLAR VA METODLAR

Tajribalar VoxCeleb1 datasetida o'tkazildi. Ushbu dataset real recording sharoitlari, turli kanal va shovqin holatlari bilan shaxsni ovozi orqali tanib olish algoritmlarini baholashda keng qo'llanadi [8].

DGSE quyidagi bosqichlardan iborat:

**1-bosqich.** Log-Mel spektrogramma hosil qilish. Kirish nutq signali  $x(t)$  dan log-Mel spektrogramma  $S(m, n)$  olinadi. Log-Mel shaxsni ovozi orqali tanib olish va nutqni qayta ishlashda samarali front-end sifatida keng ishlatiladi.

**2-bosqich.** Temporal delta hisoblash. Har bir Mel kanal bo'yicha vaqt bo'yicha o'zgarish aniqlanadi:

$$\Delta(m, n) = S(m, n) - S(m, n-1),$$

delta belgilarning foydasi klassik nutq ishlov berish adabiyotlarida ko'rsatilgan.

**3-bosqich.** Adaptiv threshold. Fiksatsiyalangan threshold o'rniga quyidagi ko'rinishdagi adaptiv threshold ishlatiladi:

$$T(m, n) = \alpha \cdot A(m, n) + \beta,$$

bu yerda  $\alpha$  – adaptiv komponent og'irligi,  $\beta$  esa bias hisoblanadi.

**4-bosqich.** Energy gate. Past energiyali va kam informativ frame larni chiqarib tashlash uchun energy gate joriy qilinadi:

$$E(n) > \theta_E.$$

**5-bosqich.** Musbat va manfiy spike kodlash. Agar temporal delta threshold dan katta bo'lsa, musbat spike hosil qilinadi:

$$P(m, n) = \begin{cases} 1, & \Delta(m, n) > T(m, n) \text{ va } E(n) > \theta_E, \\ 0, & \text{aks holda,} \end{cases}$$

agar temporal delta manfiy yo'nalishda threshold dan o'tsa, manfiy spike generatsiya qilinadi:

$$N(m, n) = \begin{cases} 1, & \Delta(m, n) < -T(m, n) \text{ va } E(n) > \theta_E, \\ 0, & \text{aks holda.} \end{cases}$$

Natijada ikki kanalli siyrak spike representation olinadi, bunday hodisaga asoslangan encoding event-driven hisoblash paradigmasi bilan mos tushadi.

- DGSE samaradorligi quyidagi mezonlar bilan baholandi:
- total spike rate – barcha pozitsiyalarga nisbatan hosil bo'lgan spike lar ulushi;
- sparsity – nol qiymatli elementlar ulushi;
- gate open rate – energy gate dan o'tgan frame lar ulushi;
- score – parametrik qidiruvda tanlash uchun ishlatilgan integrallashgan mezon.

Parametrlar uch bosqichda qidirildi:

- 1-bosqich – qo‘pol grid search;  
 2-bosqich – kengaytirilgan qidiruv;  
 3-bosqich – fine search.  
 Bu yondashuv optimal parametrga bosqichma-bosqich yaqinlashish imkonini berdi.

### 3 TAJRIBA SINOV NATIJALARI

1-bosqich natijalari – dastlabki qidiruv natijalari energy threshold noto‘g‘ri tanlanganda gate deyarli yopilib qolishini ko‘rsatdi. Natijada spike rate juda kichik, sparsity esa amaliy emas, sun‘iy ravishda juda yuqori bo‘lib qoldi.

**1-jadval.** Dastlabki qo‘pol grid search natijalari

alpha	beta	energy_thr	total_spike_rate	sparsity	gate_open_rate	score
0.8	0.01	-0.2	0.000267	0.999733	0.002167	0.438735
0.8	0.02	-0.2	0.000258	0.999742	0.002167	0.438741
1.0	0.01	-0.2	0.000177	0.999823	0.002167	0.438803
1.2	0.01	-0.2	0.000125	0.999875	0.002167	0.438827
1.4	0.01	-0.2	0.000074	0.999926	0.002167	0.438860

2-bosqich natijalari - ikkinchi bosqichda energy threshold manfiyroq diapazonga surildi va representation sezilarli yaxshilandi. Gate open rate amaliy diapazonga chiqdi, spike rate esa foydali qiymatlarga yaqinlashdi.

**2-jadval.** Kengaytirilgan qidiruv natijalari

alpha	beta	energy_thr	total_spike_rate	sparsity	gate_open_rate	score
1.0	0.05	-5.5	0.084313	0.915687	0.764533	0.037891
1.0	0.03	-5.0	0.079008	0.920992	0.666800	0.041047
1.0	0.02	-5.0	0.081600	0.918400	0.666800	0.041569
1.2	0.01	-5.5	0.070348	0.929652	0.764533	0.043537
1.0	0.01	-5.0	0.084257	0.915743	0.666800	0.044154

3-bosqich natijalari va optimal parameter - fine search bosqichida qidiruv diapazoni toraytirildi va yakuniy optimal nuqta aniqlandi.

**3-jadval.** Fine search natijalari va optimal konfiguratsiyalar

alpha	beta	energy_thr	total_spike_rate	sparsity	gate_open_rate	score
1.00	0.05	-5.25	0.079585	0.920415	0.718533	0.004230
1.05	0.03	-5.25	0.078578	0.921422	0.718533	0.005228
1.00	0.04	-5.25	0.082195	0.917805	0.718533	0.005899
1.00	0.06	-5.25	0.077012	0.922988	0.718533	0.006793
0.95	0.06	-5.25	0.083190	0.916810	0.718533	0.006829

Yakuniy optimal parametrlar: alpha = 1.0, beta = 0.05, energy\_thr = -5.25, ushbu holatda: total spike rate = 0.079585, sparsity = 0.920415, gate open rate = 0.718533 bo‘ldi.

Bosqichlar bo‘yicha eng yaxshi score ning kamayib borishi uch bosqichli qidiruvning samaradorligini ko‘rsatadi (1-rasm). Fine search bosqichida esa beta parametrining score ga ta‘siri optimal bias qiymatini aniqlashga yordam berdi (2-rasm).

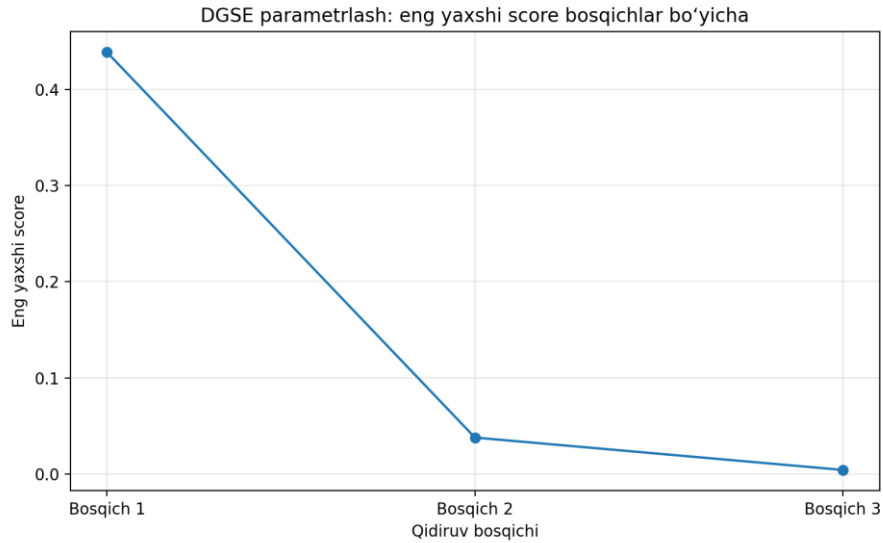
Olingan natijalar DGSE oddiy threshold kodlash emasligini ko‘rsatadi. Agar encoding faqat qattiq kesish asosida qurilganida, representation yoki haddan tashqari siyrak, yoki ortiqcha faol bo‘lib qolgan bo‘lardi. Bu ishda temporal delta, adaptiv threshold va energy gate birgalikda ishlagani sababli spike faolligi nazorat ostida saqlangan.

Birinchi bosqich natijalari juda muhim saboq berdi: katta energy threshold signalning deyarli hammasini kesib tashladi. Natijada juda yuqori sparsity hosil bo‘ldi, ammo bu foydali emas edi. Demak, representation sifati faqat sparsity bilan emas, balki informativ signal qanchalik saqlanib qolayotgani bilan ham baholanishi kerak.

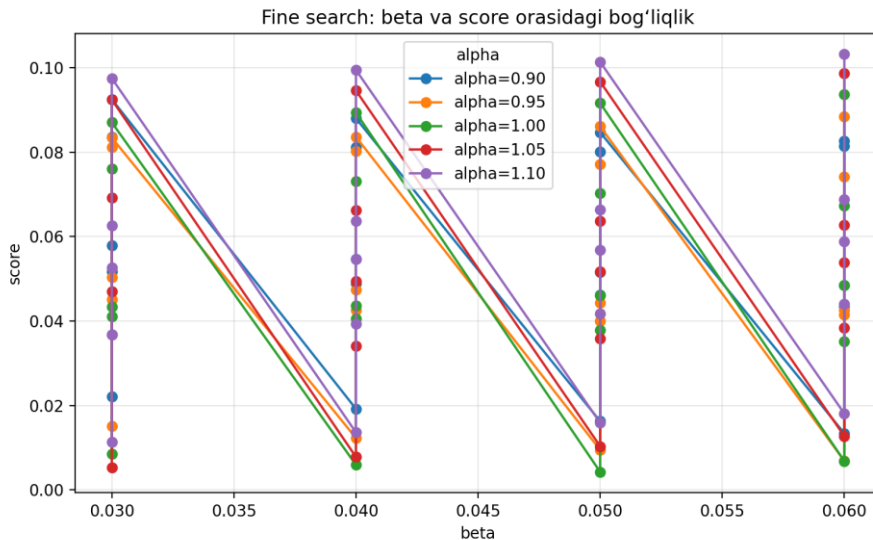
Ikkinchi va uchinchi bosqich natijalari DGSE uchun muvozanatli ish rejimini ko‘rsatdi. Yakuniy parametrlar ostida spike rate taxminan 0.08, sparsity taxminan 0.92 va gate open rate taxminan 0.72 bo‘ldi. Bu representation yetarlicha siyrak ekanini, shu bilan birga gate signalni to‘liq yopib qo‘ymaganini bildiradi.

Zamonaviy shaxsni ovozi orqali tanib olish tizimlari ko'pincha x-vector yoki ECAPA-TDNN kabi embedding arxitekturalarga tayanadi. DGSE bunday tizimlarda dense front-end o'rniga spike-oriented yoki gibrud front-end sifatida ishlatilishi mumkin, ayniqsa event-driven yoki energy-efficient hisoblash maqsad qilinganda.

Mazkur ishning cheklovi shundaki, unda hozircha asosan front-end encoding statistikasi tahlil qilindi. End-to-end speaker verification ko'rsatkichlari, masalan EER yoki minDCF, bu maqolaga kiritilmadi. Shunga qaramay, DGSE mustaqil feature extraction g'oyasi sifatida ilmiy qiymatga ega.



**1-rasm.** DGSE parametrlashda eng yaxshi scoring bosqichlar bo'yicha kamayishi



**2-rasm.** Fine search bosqichida beta va score orasidagi bog'liqlik

## 4 XULOSA

Mazkur maqolada shaxsni ovozi orqali tanib olish masalasi uchun Delta-Gated Spike Encoding (DGSE) nomli yangi feature extraction usuli taklif qilindi. Ushbu yondashuv log-Mel spektrogramma, temporal delta, adaptiv threshold, energy gate va musbat/manfiy spike encoding bosqichlarini birlashtiradi. Klassik Mel asosidagi belgilar, temporal dinamika va zamonaviy speaker representation yondashuvlari fonida DGSE shaxsni ovozi orqali tanib olish uchun siyrak va event-oriented front-end sifatida ko'rilishi mumkin.

Natijalar DGSE shaxsni ovozi orqali tanib olish uchun siyrak, nazoratlangan va amaliy spike representation hosil qilishini ko'rsatdi. Shu sababli ushbu usul resurslari cheklangan qurilmalar, SNN

asosidagi modellar va CNN-SNN gibrad arxitekturalar uchun istiqbolli front-end sifatida qaralishi mumkin. Kelgusida DGSE ni end-to-end speaker verification tizimlariga integratsiya qilib, EER, minDCF, latency va energy consumption bo'yicha to'liq baholash maqsadga muvofiq bo'ladi.

## ADABIYOTLAR

- [1] *Davis S., Mermelstein P.* Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1980. Vol. 28, No. 4. P. 357–366.
- [2] *Furui S.* Speaker-independent isolated word recognition using dynamic features of speech spectrum // IEEE Transactions on Acoustics, Speech, and Signal Processing. 1986. Vol. 34, No. 1. P. 52–59.
- [3] *Furui S.* Recent advances in speaker recognition // Pattern Recognition Letters. 1997. Vol. 18, No. 9. P. 859–872.
- [4] *Nagrani A., Chung J. S., Zisserman A.* VoxCeleb: A large-scale speaker identification dataset // Proc. Interspeech. 2017. P. 2616–2620.
- [5] *Chung J. S., Nagrani A., Zisserman A.* VoxCeleb2: Deep speaker recognition // Proc. Interspeech. 2018. P. 1086–1090.
- [6] *Snyder D., Garcia-Romero D., Sell G., Povey D., Khudanpur S.* X-vectors: Robust DNN embeddings for speaker recognition // Proc. ICASSP. 2018. P. 5329–5333.
- [7] *Desplanques B., Thienpondt J., Demuyne K.* ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in TDNN Based Speaker Verification // Proc. Interspeech. 2020. P. 3830–3834.
- [8] *Nagrani A., Chung J. S., Xie W., Zisserman A.* VoxCeleb: Large-scale speaker verification in the wild // Computer Speech & Language. 2020. Vol. 60. Article 101027.
- [9] *Rathi N. et al.* Exploring neuromorphic computing based on spiking neural networks: Algorithms to hardware // ACM Computing Surveys. 2023. Vol. 55, No. 12. P. 1–49.
- [10] *O'Shaughnessy D.* Review of methods for automatic speaker verification // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2024. Vol. 32. P. 172–198.

*Поступила в редакцию 21.01.2026*

**Citation:** Nurimov P.B., Babomuradov O.J. (2026). Delta-Gated Spike Encoding (DGSE) asosidagi siyrak va energiya-tejamkor belgilar to'plamini ajratish usuli. Raqamli texnologiyalarning nazariy va amaliy masalalari xalqaro jurnali. 9(2). – B. 86-91. <https://doi.org/10.62132/ijdt.v9i2.379>.

## A METHOD FOR EXTRACTING SPARSE AND ENERGY-EFFICIENT FEATURE SETS BASED ON DELTA-GATED SPIKE ENCODING (DGSE)

<sup>+</sup> Nurimov P.B.<sup>1</sup>, Babomuradov O.J.<sup>2</sup>

<sup>1</sup> National Research University “TIAME”, Tashkent, Uzbekistan

<sup>2</sup> Jizzakh Branch of Kazan Federal University, Jizzakh, Uzbekistan

**Abstract.** This paper proposes a new feature extraction method for speaker recognition, called Delta-Gated Spike Encoding (DGSE). The proposed approach combines log-Mel spectrograms, temporal delta, adaptive thresholding, an energy gate, and positive/negative spike encoding stages. The goal is to extract informative time-frequency changes from the acoustic signal and form a sparse representation suitable for subsequent spiking or energy-efficient models. Experiments were conducted on the VoxCeleb1 dataset using a three-stage parametric search procedure. The results of the initial coarse search, the subsequent extended search, and the final fine search were compared. The best result was obtained with the parameters  $\alpha = 1.0$ ,  $\beta = 0.05$ , and  $\text{energy\_thr} = -5.25$ . Under these settings, the method achieved total spike rate = 0.079585, sparsity = 0.920415, and gate open rate = 0.718533. The obtained results show that the DGSE method preserves informative parts of the signal while ensuring high sparsity. This makes it a promising solution for resource-constrained devices and CNN-SNN hybrid speaker recognition systems.

**Keywords:** speaker recognition, feature extraction, log-Mel, spike encoding, sparsity, neuromorphic computing, VoxCeleb1, DGSE.

## МЕТОД ИЗВЛЕЧЕНИЯ РАЗРЕЖЕННЫХ И ЭНЕРГОЭФФЕКТИВНЫХ ПРИЗНАКОВ НА ОСНОВЕ DELTA-GATED SPIKE ENCODING (DGSE)

<sup>+</sup> *Нуримов П.Б.<sup>1</sup>, Бабомурадов О.Ж.<sup>2</sup>*

<sup>1</sup> Национальный исследовательский университет «ТИИИМСХ»,  
Ташкент, Узбекистан

<sup>2</sup> Джизакский филиал Казанского федерального университета,  
Джизак, Узбекистан

**Аннотация.** В данной статье предлагается новый метод извлечения признаков для задачи распознавания личности по голосу – Delta-Gated Spike Encoding (DGSE). Предложенный подход объединяет логарифмические Mel-спектрограммы, временные дельта-признаки, адаптивный порог, энергетический шлюз и этапы кодирования положительных/отрицательных спайков. Цель метода заключается в извлечении информативных временно-частотных изменений в акустическом сигнале и формировании разреженного представления, удобного для последующих спайковых или энергоэффективных моделей. Эксперименты были проведены на наборе данных VoxCeleb1 с использованием трехэтапного параметрического поиска. Были сопоставлены результаты первоначального грубого поиска, последующего расширенного поиска и финального точного поиска. Наилучший результат был получен при параметрах  $\alpha = 1.0$ ,  $\beta = 0.05$ ,  $\text{energy\_thr} = -5.25$ . В этом случае были достигнуты следующие значения:  $\text{total spike rate} = 0.079585$ ,  $\text{sparsity} = 0.920415$ ,  $\text{gate open rate} = 0.718533$ . Полученные результаты показывают, что метод DGSE сохраняет информативные части сигнала при обеспечении высокой разреженности. Это делает его перспективным решением для устройств с ограниченными ресурсами, а также для гибридных систем распознавания говорящего на основе CNN-SNN.

**Ключевые слова:** распознавание говорящего, извлечение признаков, log-Mel, spike-кодирование, разреженность, нейроморфные вычисления, VoxCeleb1, DGSE.