Daugiakalbio šnekos modelio atpažinimo gerinimas lietuvių kalbai

Juras Ezerskis, Asta Slotkienė

Vilniaus universitetas, Matematikos ir informatikos fakultetas,
Didlaukio g. 47, LT-08303 Vilnius, Lietuva
juras.ezerskis@mif.stud.vu.lt

Santrauka. Lietuvių spontaninės kalbos automatinis atpažinimas išlieka sudėtinga užduotis dėl ribotų išteklių, tarimo variacijų ir nuostolingo garso įrašų poveikio transkribavimo kokybei. Šio straipsnio tikslas – ištirti *Whisper large-v3* modelio pritaikymą lietuvių spontaninės kalbos atpažinimui taikant *Low-Rank Adaptation* (LoRA) strategiją, palyginti jį su baziniu modeliu ir išanalizuoti, kaip papildomas apmokymas veikia transkribavimo kokybę. Gauti rezultatai parodė, kad LoRA strategija reikšmingai pagerino modelio veikimą ir sumažino tiek tekstinį, tiek fonetinį nutolimą nuo tikrųjų transkripcijų.

Raktiniai žodžiai: automatinis šnekos atpažinimas, lietuvių kalba, LoRA, papildomas apmokymas.

1 Įvadas

Automatinis šnekos atpažinimas pastaraisiais metais sparčiai pažengė dėl didelių apmokytų neuroninių modelių atsiradimo. Vienas žinomiausių tokio tipo modelių yra *Whisper*, apmokytas didelės apimties daugiakalbiais garso ir teksto duomenimis, todėl galintis būti taikomas ir kalboms, kurioms trūksta lokalių išteklių [1].

Tačiau mažų išteklių kalbų atveju vien bazinio daugiakalbio modelio dažnai nepakanka. Lietuvių kalbai būdinga sudėtinga morfologinė struktūra, didelė tarimo variacija, spontaninės kalbos netolygumas ir ribotas viešai prieinamų, kokybiškai anotuotų šnekos duomenų kiekis. Dėl to net stiprūs daugiakalbiai modeliai realistiškose situacijose vis dar daro reikšmingą kiekį klaidų.

Ši problema ypač išryškėja spontaninės kalbos scenarijuje, kur pasitaiko nebaigtų sakinių, tarminių ar individualių tarimo ypatumų, akustinių trukdžių bei nuostolingo garso suspaudimo. Naujesni lietuvių kalbos tyrimai rodo, kad transkribavimo kokybę reikšmingai veikia įrašymo sąlygos, garso formatas ir paties duomenų rinkinio pobūdis [2, 3]. Todėl aktualu tirti ne tik bazinių modelių veikimą, bet ir jų pritaikymo strategijas konkrečiai lietuvių kalbai.

DOI: https://doi.org/10.15388/LMITT.2026.8

Vienas iš praktiškų būdų pritaikyti didelį modelį konkrečiai užduočiai yra parametrų požiūriu efektyvus papildomas apmokymas. LoRA strategija šiame darbe pasirinkta remiantis naujesniais mažų išteklių kalbų šnekos atpažinimo tyrimais, kuriuose parametrų požiūriu efektyvūs papildomo apmokymo metodai buvo vertinami kaip praktiška alternatyva pilnam perapmokymui ribotų duomenų ir skaičiavimo išteklių sąlygomis. Kadangi tokiuose lyginamuosiuose tyrimuose LoRA buvo nagrinėjama tarp konkurencingų adaptacijos strategijų, šiame darbe ji pasirinkta kaip pagrįsta bazinė didelio daugiakalbio modelio pritaikymo schema lietuvių spontaninės kalbos duomenims [4].

Šio straipsnio tikslas – įvertinti, kaip LoRA strategija veikia *Whisper large-v3* modelio pritaikymą lietuvių spontaninės kalbos transkribavimui, ir aptarti, kaip pasirinkta adaptacija prisideda prie transkribavimo kokybės gerinimo.

2 Susiję darbai

Mažų išteklių kalbų šnekos atpažinimo tyrimuose dažnai pabrėžiama, kad modelių veikimo kokybė priklauso ne tik nuo architektūros, bet ir nuo duomenų apimties, jų įvairovės bei atitikimo realioms taikymo sąlygoms. Dalis viešai prieinamų rinkinių apima daugiausia perskaitytą kalbą, todėl modeliai, gerai veikiantys kontroliuojamose sąlygose, gali reikšmingai prasčiau veikti spontaninės kalbos scenarijuje [5].

Whisper modelis mažų išteklių kalbų tyrimuose yra svarbus tuo, kad net ir be papildomo mokymo dažnai pasiekia konkurencingus rezultatus. Vis dėlto literatūroje matyti, kad šio modelio veikimą riboja kalbos specifika, domeno neatitiktis ir akustinės sąlygos, todėl papildomas pritaikymas išlieka aktualus [1, 6].

Papildomo apmokymo strategijų kontekste vis daugiau dėmesio skiriama *parameter-efficient fine-tuning* (PEFT) metodams, kai papildomai apmokoma tik dalis modelio parametrų, o likusi bazinio modelio dalis paliekama nekeičiama. Pilnas modelio perapmokymas dažnai reikalauja didelių GPU resursų, ilgina mokymo laiką ir didina persimokymo riziką, ypač kai mokymo duomenų kiekis ribotas. Dėl to LoRA ir kiti PEFT metodai tampa praktiškai patraukliomis alternatyvomis [7, 8, 9].

Šiame straipsnyje sąmoningai netiriamos kelios skirtingos PEFT strategijos. Vietoj to susitelkiama į vieną aiškiai apibrėžtą scenarijų: bazinio *Whisper large-v3* ir to paties modelio po LoRA papildomo apmokymo palyginimą lietuvių spontaninės kalbos duomenyse.

3 Tyrimo eiga

Tyrimą sudaro du etapai, kuriais siekiama įvertinti LoRA strategijos įtaką nekeičiant bazinio modelio architektūros ir taikant tuos pačius kiekybinius vertinimo kriterijus. Tyrimų struktūra apibendrinta 1 lentelėje.

1 lentelė. Tyrimų dizainas.

Tyrimas	Tikslas	Aprašymas
1	Bazinio modelio vertinimas	Naudojamas *openai/whisper-large-v3* modelis be papildomo apmokymo. Modelis tiesiogiai vertinamas nuostolingos lietuvių spontaninės kalbos testavimo aibėje.
2	LoRA adaptacijos vertinimas	Tas pats *Whisper large-v3* modelis papildomai apmokomas taikant LoRA strategiją ir po to vertinamas toje pačioje testavimo aibėje.

Tyrime naudotas ne visas LIEPA-2 garsynas, o jo pagrindu sudarytas spontaninės nuostolingos lietuvių kalbos poaibis [10, 11]. LIEPA-2 apima anotuotus lietuviškos šnekos įrašus, kurių techninės charakteristikos yra *.wav*, 16 kHz diskretizavimo dažnis, 16 bitų garso gylis ir vienas kanalas, o failų pavadinimuose užkoduoti metaduomenys leidžia atsirinkti skirtingo tipo įrašus. Šiame darbe buvo atrenkami tik tie įrašai, kurių pavadinimai atitiko *L_S** šabloną, t. y. nuostolingo formato (*L, lossy*) ir spontaninės kalbos (*S, spontaneous*) kriterijus. Nuostolingas formatas šiame tyrime reiškia garso suspaudimą, kai dalis akustinės informacijos prarandama. Į analizę nebuvo įtraukti nei perskaitytos kalbos, nei nenuostolingo formato įrašai. Pagal šiuos kriterijus atrinkti 122 įrašai

Duomenų paruošimo metu iš atrinktų įrašų anotacijų ir transkripcijų buvo išskirti šnekos segmentai, o tekstas normalizuotas: paverstas mažosiomis raidėmis, pašalinti skyrybos ženklai, anotaciniai intarpai ir pertekliniai tarpai. Toliau segmentai filtruoti pagal trukmę ir prireikus jungiami į ilgesnius vientisus segmentus. Prieš jungimą gauti 47 544 segmentai, po jungimo liko 23 428 segmentai. Paruošti duomenys įrašų lygmeniu suskirstyti į mokymo, validacijos ir testavimo aibes santykiu 80/10/10, užtikrinant, kad to paties įrašo segmentai nepatektų į skirtingas aibes. Galutinai sudaryta 18 844 segmentų mokymo aibė, 2 533 segmentų validacijos aibė ir 2 051 segmento testavimo aibė. Duomenų paruošimo santrauka pateikta 2 lentelėje.

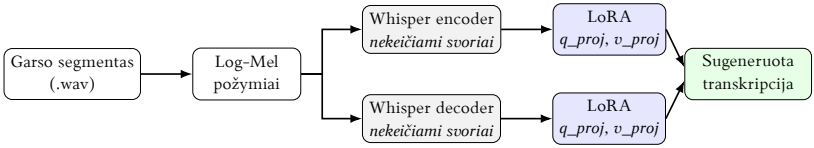
2 lentelė. Duomenų paruošimo santrauka.

Rodiklis	Reikšmė
Atrinktų įrašų skaičius	122
Segmentų skaičius prieš jungimą	47 544
Segmentų skaičius po jungimo	23 428
Mokymo aibė	18 844
Validacijos aibė	2 533
Testavimo aibė	2 051
Minimalus segmento ilgis	0.4 s
Maksimalus segmento ilgis	30 s
Tikslinė sujungto segmento trukmė	25 s
Maksimalus tarpas jungimui	0.6 s

Tyrime naudotas *openai/whisper-large-v3* modelis. Bazinė modelio architektūra nebuvo keičiama: papildomas apmokymas vykdytas taikant LoRA modulius *q_proj* ir *v_proj* linijiniams sluoksniams. Dėmesio mechanizme *q_proj* sluoksnis formuoja užklausų (*query*) reprezentacijas, pagal kurias nustatoma, į kuriuos akustinius ir kalbinius požymius modelis turėtų kreipti daugiau dėmesio, o *v_proj* sluoksnis formuoja reikšmių (*value*) reprezentacijas, iš kurių atrinkta informacija perduodama tolesniam apdorojimui. LoRA pasirinkta todėl, kad ji leidžia adaptuoti tik ribotą parametrų dalį ir taip vertinti būtent šios adaptacijos indėlį į transkribavimo kokybės gerinimą, kartu išsaugant bazinio modelio bendrąsias šnekos atpažinimo žinias.

LoRA pagrindu buvo adaptuojama tik ribota *Whisper large-v3* parametrų dalis: prie *q_proj* ir *v_proj* sluoksnių prijungti papildomi treniruojami žemo rango moduliai, o visi bazinio modelio svoriai palikti nekeičiami. Todėl adaptacija buvo sutelkta į dėmesio mechanizmą, nekeičiant viso modelio. Kadangi šie sluoksniai priklauso dėmesio mechanizmui, tikėtina, kad jų adaptavimas padėjo modeliui tiksliau susieti akustinius požymius su lietuviškais žodžių vienetais, ypač esant tarimo variacijoms ir nuostolingo garso iškraipymams.

Prieš pateikiant garsą modeliui, garso signalas paverčiamas į logaritminės Mel spektrogramos požymius (*log-Mel features*), kurie naudojami kaip *Whisper* įvestis. Taikant pasirinktą LoRA adaptacijos schemą, baziniai *Whisper* encoderio ir decoderio svoriai išlieka nekeičiami, o papildomai treniruojami moduliai prijungiami tik prie dėmesio mechanizmo projekcijų *q_proj* ir *v_proj*. Tokiu būdu adaptuojama tik ribota modelio parametrų dalis, išlaikant bazinio modelio žinias. Ši apdorojimo ir adaptavimo schema parodyta 1 paveiksle.



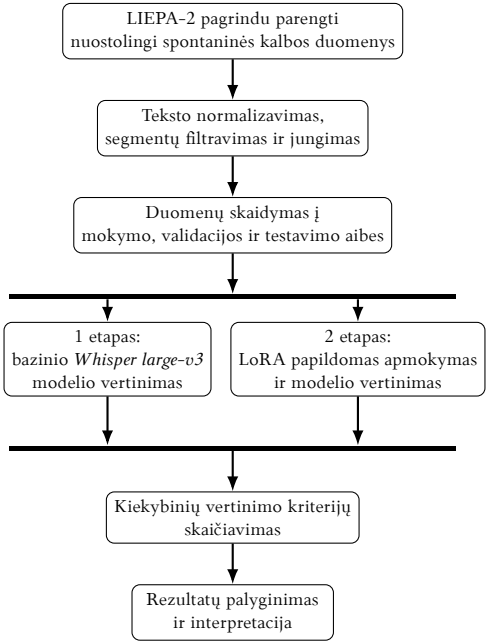
1 pav. Whisper large-v3 architektūros pritaikymo schema taikant LoRA.

Papildomo apmokymo bazinė LoRA konfigūracija šiame darbe parinkta remiantis artimais mažų išteklių kalbų šnekos atpažinimo tyrimais. Hsieh ir kt. (2025), taikydami *Whis-per*+LoRA taivaniečių hakka kalbai, naudojo konfigūraciją r = 16, α = 32 ir *dropout*=0.05, todėl šios reikšmės šiame darbe pasirinktos kaip literatūroje pagrįstas bazinis parametrų derinys. Papildomai, Kim ir kt. (2023) parodė, kad LoRA taikymas dėmesio sluoksniams yra perspektyvi kryptis mažų išteklių kalbų atveju, o Rafkin ir kt. (2026) mažų išteklių *Whisper-large-v3* tyrime taikė LoRA modulius *q_proj* ir *v_proj* sluoksniams. Todėl 3 lentelėje pateikta konfigūracija šiame darbe traktuojama kaip literatūra paremta bazinė LoRA adaptacijos schema, o ne kaip išsamiai optimizuotas hiperparametrų derinys [12, 13, 14].

3 lentelė. Modelio ir papildomo apmokymo hiperparametrai.

Parametras	Reikšmė
Bazinis modelis	*openai/whisper-large-v3*
Kalba / užduotis	*lithuanian* / *transcribe*
Tikslinis diskretizavimo dažnis	16 kHz
LoRA tiksliniai sluoksniai	*q_proj*, *v_proj*
LoRA rangas r	16
LoRA α	32
LoRA *dropout*	0.05
Epochų skaičius	3
Mokymosi greitis	10^{-4}
Warmup ratio	0.05
Batch size (train / eval)	1 / 1
Gradient accumulation steps	8
Maksimalus išvesties ilgis	225
Modelio svorių kvantizacija	8-bit
Mišrus tikslumas	FP16, jei prieinama CUDA
Optimizatorius	*paged_adamw_8bit*

Siekiant aiškiai parodyti viso tyrimo eigą nuo duomenų paruošimo iki rezultatų interpretacijos, 2 paveiksle pateikta tyrimo etapų veiklos diagrama.



2 pav. Tyrimo etapų veiklos diagrama.

Segmentų lygmens prognozės ir iš jų apskaičiuoti vertinimo duomenys pateikti GitHub saugykloje[1]. Modelių veikimas vertintas pagal penkis kiekybinius vertinimo kriterijus: žodžių klaidų dažnį (WER), simbolių klaidų dažnį (CER), semantinį panašumą (SEM), raktažodžių atkūrimo rodiklį (KR) ir fonemų redagavimo atstumą (PED). WER ir CER naudoti paviršiniam tekstiniam tikslumui vertinti, SEM – sugeneruoto teksto prasminiam artumui tikrajai transkripcijai įvertinti, KR – svarbiausių žodžių išsaugojimui nustatyti, o PED – fonetiniam nutolimui tarp tikrojo teksto ir prognozės vertinti. WER ir CER yra plačiai taikomi automatinio šnekos atpažinimo vertinimo matai, tačiau naujesni darbai pabrėžia, kad vien jų nepakanka semantiniam ir fonetiniam prognozių artumui įvertinti [15, 16].

[1] https://github.com/JurasEz/Whisper-Model-Fine-Tuning/tree/main/results

4 Rezultatai

Galutinis bazinio ir LoRA pagrindu papildomai apmokyto modelio palyginimas pateikiamas 4 lentelėje. Palyginimas atliktas naudojant visą testavimo aibę, sudarytą iš 2 051 segmento.

4 lentelė. Bazinio ir LoRA pagrindu papildomai apmokyto modelio palyginimas.

Metrika	Bazinis	LoRA	Absoliutus pokytis	Santykinis pokytis
WER	0.3470	0.2116	-0.1354	-39.0%
CER	0.1540	0.0577	-0.0963	-62.5%
SEM	0.8417	0.9219	+0.0802	+9.5%
KR	0.5490	0.7465	+0.1975	+36.0%
PED	0.2876	0.0681	-0.2195	-76.3%

4 lentelės rezultatai rodo, kad LoRA pagrindu papildomai apmokytas *Whisper large-v3* modelis pranoko bazinį modelį visose vertintose metrikose. WER sumažėjo nuo 0.3470 iki 0.2116, o CER – nuo 0.1540 iki 0.0577, todėl galima teigti, kad papildomas apmokymas pagerino tiek žodžių sekų, tiek simbolinės struktūros atkūrimą.

Pagerėjo ir papildomos kokybės metrikos: SEM padidėjo nuo 0.8417 iki 0.9219, o KR – nuo 0.5490 iki 0.7465. Tai rodo, kad papildomai apmokytas modelis ne tik sumažino paviršinių klaidų skaičių, bet ir geriau išsaugojo sakinio prasmę bei svarbiausią leksinę informaciją.

Tikėtina, kad pagerėjimą lėmė ne vien pats papildomas apmokymas, bet ir konkreti pasirinkta LoRA adaptacijos schema. Šiame tyrime LoRA moduliai buvo taikyti *q_proj* ir *v_proj* sluoksniams, tiesiogiai susijusiems su dėmesio mechanizmu, o bazinio *Whisper large-v3* modelio svoriai palikti nekeičiami. Tokia konfigūracija leido adaptuoti, kaip modelis formuoja ir paskirsto dėmesį akustiniams bei kalbiniams požymiams, todėl jis galėjo tiksliau išskirti svarbiausius spontaninės lietuvių kalbos informacinius vienetus. Tikėtina, kad gautiems rezultatams įtakos turėjo ir pasirinkta LoRA hiperparametrų konfigūracija, tačiau šiame darbe nebuvo siekiama atskirai įvertinti skirtingų hiperparametrų reikšmių poveikio. Tai gali paaiškinti ne tik mažesnį WER ir CER, bet ir SEM bei KR pagerėjimą, rodantį geresnį prasmės ir svarbiausios leksinės informacijos išsaugojimą.

PED rezultatai taip pat rodo aiškų pagerėjimą: bazinio modelio reikšmė buvo 0.2876, o LoRA modelio – 0.0681. Kadangi mažesnė PED reikšmė reiš-

kia mažesnį fonetinį nutolimą nuo tikrosios transkripcijos, galima teigti, kad LoRA modelio prognozės tapo artimesnės teisingoms ne tik tekstiniu, bet ir fonetiniu požiūriu.

Papildoma analizė parodė, kad testavimo metu abu modeliai sėkmingai sugeneravo prognozes visiems segmentams, t. y. nebuvo užfiksuota techninių vykdymo klaidų. Tikslaus sutapimo dažnis padidėjo nuo 0.1438 iki 0.3345, o tai rodo, kad LoRA modelis gerokai dažniau visiškai tiksliai atkūrė tikrąją transkripciją. Abiem modeliams KR aprėptis siekė 0.978, todėl raktažodžių atkūrimo metrikos palyginimas buvo atliktas nuosekliomis sąlygomis.

5 Rezultatų aptarimas

Šiame darbe lyginami ne keli skirtingi PEFT metodai, o vienas aiškiai apibrėžtas scenarijus: bazinis modelis ir tas pats modelis po LoRA papildomo apmokymo. Toks tyrimo dizainas leidžia tiesiogiai įvertinti, ar LoRA strategija naudinga lietuvių spontaninės kalbos atpažinimo užduočiai.

LoRA strategija šiame tyrime pasirodė naudinga lietuvių spontaninės kalbos atpažinimo užduočiai. Tikėtina, kad šį pagerėjimą lėmė tai, jog LoRA adaptacija buvo taikyta dėmesio mechanizmo projekcijoms *q_proj* ir *v_proj*. Kadangi šie sluoksniai dalyvauja paskirstant dėmesį akustiniams ir kalbiniams požymiams, jų pritaikymas galėjo padėti modeliui tiksliau apdoroti spontaninės lietuvių kalbos tarimo, ritmo ir fonetinio kintamumo ypatumus. Tai ypač svarbu spontaninės kalbos kontekste, nes vien paviršinis teksto sutapimas ne visada pilnai atspindi transkripcijos kokybę. Vis dėlto šiame darbe rezultatai gauti vienoje duomenų skaidymo schemoje, todėl nebuvo atskirai vertintas jų stabilumas taikant pakartotinius bandymus ar kryžminę validaciją. Dėl to gauti skirtumai tarp bazinio ir LoRA modelio šiame darbe interpretuojami kaip šiame tyrimo nustatyme stebėta tendencija, o ne kaip galutinai apibendrinta išvada apie visus galimus lietuvių spontaninės kalbos duomenų atvejus. Todėl LoRA strategija šiuo atveju gali būti laikoma perspektyvia didelio daugiakalbio modelio adaptavimo kryptimi, neišplečiant tyrimo iki pilno visų parametrų perapmokymo.

Šiame straipsnio variante dar nepateikiami detalūs grafikai pagal amžiaus grupes, įrašo tipą, lytį ar nuostolingumo pobūdį, nors tokia analizė buvo apskaičiuota.

6 Išvados

Kaip parodė 4 lentelėje pateikti rezultatai, papildomai apmokytas modelis šiame tyrime pasiekė mažesnį žodžių ir simbolių klaidų dažnį, didesnį semantinį panašumą, geresnį raktažodžių atkūrimą ir mažesnį fonetinį nutolimą nuo tikrųjų transkripcijų. Tai rodo, kad LoRA adaptacija šiame tyrimo nustatyme buvo susijusi su geresniais modelio veikimo rodikliais, lyginant su baziniu modeliu. Papildomi rodikliai, tokie kaip padidėjęs tikslaus sutapimo dažnis ir stabili KR aprėptis, taip pat patvirtina, kad stebėtas pagerėjimas nebuvo susijęs su techniniais vykdymo sutrikimais.

Gauti rezultatai leidžia sieti šį pagerėjimą ne vien su pačiu papildomu apmokymu, bet ir su konkrečia LoRA taikymo schema: moduliai buvo integruoti į *q_proj* ir *v_proj* sluoksnius, susijusius su dėmesio mechanizmu, o bazinio modelio svoriai palikti nekeičiami. Taigi šis tyrimas patvirtina, kad LoRA gali būti veiksmingas būdas pritaikyti didelį daugiakalbį modelį lietuvių kalbai, kaip mažų išteklių kalbos scenarijui.

Vis dėlto šiame darbe nebuvo siekiama nustatyti, kuris LoRA sluoksnių ir hiperparametrų derinys yra geriausias, todėl skirtingų parametrizavimo sprendimų poveikis turėtų būti tiriamas tolimesniuose tyrimuose. Tolimesniuose tyrimuose būtų tikslinga analizuoti skirtingų sluoksnių ir hiperparametrų įtaką, siekiant nustatyti konfigūraciją, kuri leistų pasiekti dar didesnį pagerėjimą.

Literatūra

[1] A. Radford ir kiti. „Robust Speech Recognition via Large-Scale Weak Supervision“. Iš: (2022).

[2] J. Šablevičius, A. Slotkienė. „The Impact of Audio Formats, Recording Environments, and Playback Speeds on the Accuracy of Lithuanian Speech Recognition“. Iš: *Proceedings of the 25th International Conference on Speech and Computer (SPECOM)*. 2025.

[3] P. Kasparaitis. „Evaluation of Lithuanian Speech-to-Text Transcribers“. Iš: *Informatica* 36.2 (2025), puslapiai 369–384.

[4] S. Imam ir kiti. „Full Fine-Tuning vs. Parameter-Efficient Adaptation for Low-Resource African Speech Recognition“. Iš: *Proceedings of the 7th Workshop on African Natural Language Processing*. 2026. URL: https://aclanthology.org/2026.africanlp-main.19.pdf.

[5] K. Micallef ir kiti. „MASRI-HEADSET: A Maltese Corpus for Speech Recognition“. Iš: *arXiv preprint arXiv:2306.01009* (2023).

[6] Y. Liu. „Exploring Whisper Fine-Tuning Strategies for Low-Resource Speech Recognition“. Iš: *EURASIP Journal on Audio, Speech, and Music Processing* (2024).

[7] J. Mainzinger. „Fine-Tuning ASR Models for Very Low-Resource Languages“. Iš: *ACL Workshop on Speech and Language Technologies*. 2024.

[8] L. G. Pillai. „Multistage Fine-Tuning Strategies for Automatic Speech Recognition in Low-Resource Languages“. Iš: (2024).
[9] H. Wang. „Low-Resource Speech Recognition by Fine-Tuning Whisper with LoRA“. Iš: *Applied Sciences* (2025).
[10] *Apie projektą LIEPA-2*. Raštija.lt. URL: https: //xn--ratija-ckb.lt/liepa-2/apie-projekta-liepa-2/.
[11] *Garsynas LIEPA-2*. Raštija.lt. URL: https: // xn-- atija-ckb.lt/liepa-2 /infrastrukturines-paslaugos/garsynas/.
[12] H. Hsieh ir kiti. „A Compact Whisper+LoRA Baseline for Taiwanese Hakka“. Iš: *Proceedings of ROCLING 2025*. 2025. URL: https://aclanthology.org/2025.rocling-main.55.pdf.
[13] H.-S. Kim ir kiti. „Adapt and Prune Strategy for Multilingual Speech Recognition in Low-Resource Settings“. Iš: *Proceedings of the Workshop on Multilingual Representation Learning*. 2023. URL: https://aclanthology.org/2023.mrl-1.7.pdf.
[14] E. Rafkin ir kiti. „Task Arithmetic with Support Languages for Low-Resource ASR“. Iš: *arXiv preprint arXiv:2601.07038* (2026). URL: https://arxiv.org/pdf/2601. 07038.
[15] R. Ellis, W. Byrne, S. Shanmugam ir kiti. „WER Is Unaware: Assessing How ASR Errors Distort Clinical Understanding in Patient-Facing Dialogue“. Iš: *arXiv preprint arXiv:2511.16544* (2025).
[16] B. Phukon, X. Zheng, M. Hasegawa-Johnson. „Aligning ASR Evaluation with Human and LLM Judgments: Intelligibility Metrics Using Phonetic, Semantic, and NLI Approaches“. Iš: *arXiv preprint arXiv:2506.16528* (2025).