

Giliojo mokymo modeliai kalbos signalui gryninti

Parengė: Daniel Zakševski

Darbo vadovas: Doc. Dr. Gintautas Tamulevičius

Doktorantūros studijų metai: 2023-2027

Studijų metai: 1-ieji

Doktorantūros studijų ir mokslinių tyrimų plano bei jo vykdymo suvestinė

Studijų metai	Egzaminai	
	Planas	Įvykdyta
I (2023/2024)	2	2
II (2024/2025)	1	
III (2025/2026)	1	
IV (2026/2027)		
Iš viso:	4	2

Studijų metai	Dalyvavimas konferencijose				Publikacijos					
	Tarptautinėse		Nacionalinėse		Su citavimo rodikliu			Be citavimo rodiklio		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė	Planas	Įvykdyta	Būklė
I (2023/2024)	0		0		0			0		
II (2024/2025)	0		0		0			1		
III (2025/2026)	1		1		1			0		
IV (2026/2027)	2		1		1			0		
Iš viso:	3		2		2			1		

Ataskaitinio pusmečio darbo planas ir jo vykdymas, 2023/2024 – I pusmetis

Egzaminai, 2023/2024 – II pusmetis		
Planas	Ivykdyta	Būklė
Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika (2024.06.25)	Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika, 2024.06.25	Išlaikyta

Dalyvavimas konferencijose, 2023/2024 – II pusmetis		
Planas	Ivykdyta	Konferencijos tipas

Publikacijos, 2023/2024 – II pusmetis			
Planas	Ivykdyta	Būklė	Publikacijos tipas

Doktorantūros studijų pasiekimai, tarptautiniai renginiai ir publikacijos

Dalyvavimas tarptautinėse konferencijose	
	Aprašas

Publikacijos su citavimo rodikliu	
	Bibliografinis aprašas

Mokslinių tyrimų ir disertacijos rengimo etapai (1)

Darbo pavadinimas		Atlikimo terminai		Pastabos
		Nuo	Iki	
1.	Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje):			1.1., 1.2. įvykdyti, 1.3, 1.4 vykdomi
1.1.	Kalbos signalo gryninimo problematikos apžvalga	2023 m. IV ketvirtis	2024 m. I ketvirtis	
1.2.	Kalbos signalo gryninimo problemos klasikinių sprendimų metodikos analizė	2024 m. I ketvirtis	2024 m. I ketvirtis	
1.3.	Kalbos signalo gryninimo problemos mašininio mokymu grįstų sprendimų metodikos analizė	2024 m. I ketvirtis	2024 m. II ketvirtis	
1.4.	Išankstinio duomenų apdorojimo metodų analizė	2024 m. I ketvirtis	2024 m. III ketvirtis	
2.	Mokslinio tyrimo vykdymas:			2.1.1 pradėtas vykdyti
2.1.	Tyrimo metodikos sudarymas:			
2.1.1.	Tyrimų tikslų ir užduočių suformulavimas, tyrimo metodikos sudarymas	2024 m. II ketvirtis	2024 m. III ketvirtis	
2.2.	Teorinis tyrimas:			
2.2.1.	Klasikinių bei mašininio mokymu grįstų metodų palyginimas kalbos signalo gryninimo uždavinio sprendimui	2024 m. III ketvirtis	2025 m. I ketvirtis	
2.2.2.	Giliojo mokymo metodais grįstų kalbos signalo gryninimo algoritmų apibendrinimas	2025 m. I ketvirtis	2025 m. III ketvirtis	

Mokslinių tyrimų ir disertacijos rengimo etapai (2)

Darbo pavadinimas		Atlikimo terminai		Pastabos
		Nuo	Iki	
2.3.	Empirinis tyrimas:			
2.3.1.	Skirtingomis architektūromis paremtų giliojo mokymo algoritmų tyrimas	2025 m. III ketvirtis	2026 m. I ketvirtis	
2.3.2.	Giliojo mokymo metodais grįsto kalbos gryninimo algoritmų praktinio taikymo problematikos tyrimas	2026 m. I ketvirtis	2026 m. IV ketvirtis	
2.4.	Gautų duomenų analizė, apibendrinimas, išvadų parengimas:			
2.4.1.	Gautų duomenų analizė, apibendrinimas, išvadų parengimas	2026 m. IV ketvirtis	2027 m. III ketvirtis	
3.	Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų, ir kt.) parengimas:			
3.1.	Literatūros apžvalga	2026 m. IV ketvirtis	2027 m. I ketvirtis	
3.2.	Tyrimų metodikos aprašymas	2027 m. I ketvirtis	2027 m. II ketvirtis	
3.3.	Rezultatų aprašymas	2027 m. II ketvirtis	2027 m. III ketvirtis	
3.4.	Ginamų teiginių ir išvadų aprašymas	2027 m. III ketvirtis	2027 m. IV ketvirtis	
4.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2027 m. IV ketvirtis	2027 m. IV ketvirtis	
5.	Daktaro disertacijos gynimas	2027 m. IV ketvirtis	2027 m. IV ketvirtis	

Tyrimo detalės

- Tyrimo objektas: šnekos signalas, balso signalo kokybė bei triukšmas.
- Tyrimo tikslai: formuluojami.
- Tyrimo uždaviniai: gryninami.

Moksliniai rezultatai, 2023/2024 – II pusmetis (1)

- Pagrindinis kalbos gryninimo (angl. speech enhancement) tikslas – padidinti akustinių kalbos signalų kokybę ir suprantamumą, pašalinant įvairius triukšmus bei jų sukeltus iškraipymus, atkuriant ar išryškinant spektrines kalbos signalo savybes.
- Natūralaus kalbos gryninimo proceso pavyzdys – Lombardo efektas. Kalbėdamas triukšmingoje aplinkoje, žmogus pakelia toną, pabrėžtinai taria garsus, dėl ko šnekos signalas tampa lengviau išskiriamas triukšme ir suprantamas.

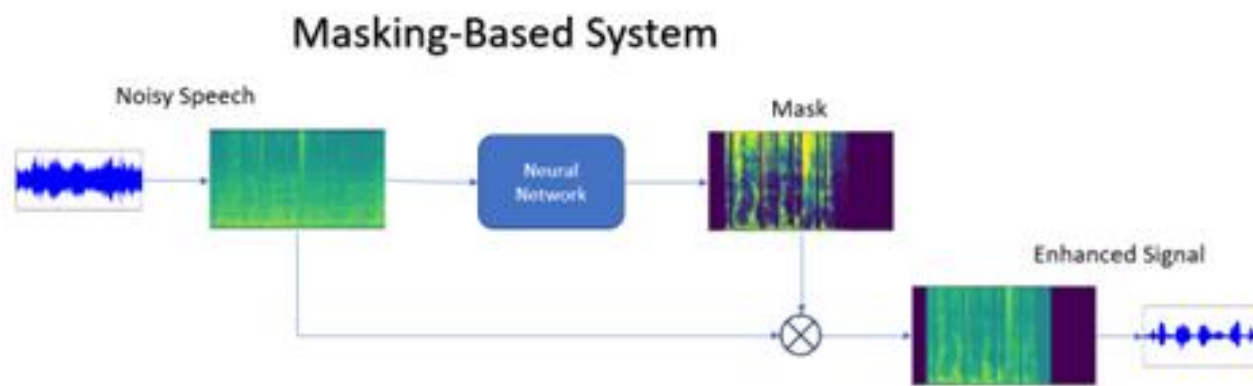
Moksliniai rezultatai, 2023/2024 – II pusmetis (2)

- Kalbos apdorojimo srityje iki dirbtinio intelekto metodų įsigalėjimo klasikiniai skaitmeninio signalo apdorojimo metodai.
- Šie metodai reikalauja išankstinių žinių apie triukšmą arba akustinę aplinką.
- Privalumai: taikymo paprastumas bei mažas kompiuterinių resursų naudojimas, duomenų poreikis – minimalus.
- Trūkumai: Yra itin jautrūs triukšmo savybių pokyčiams.

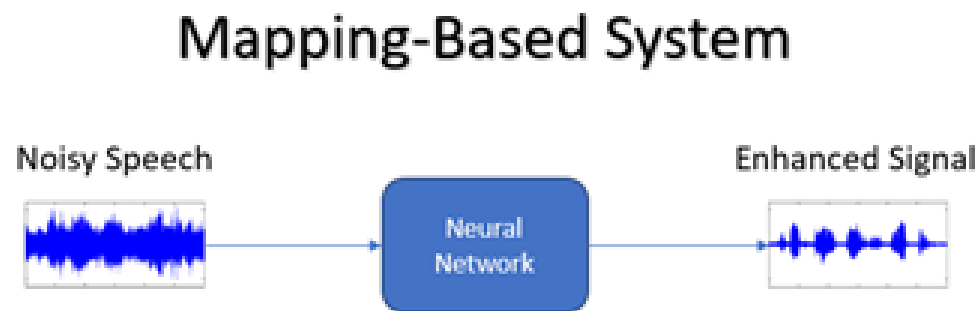
Moksliniai rezultatai, 2023/2024 – II pusmetis (3)

- Mašininio mokymo sprendimai, ypač paremti giliuoju mokymu, šiuo metu yra vyraujanti technologija sprendžiant kalbos gryninimo problemą.
- Galima išskirti du dažniausiai literatūroje sutinkamus algoritmų veikimo mechanizmus; jie gali būti paremti maskavimu (angl. *masking*) arba susiejimu (angl. *mapping*).

Moksliniai rezultatai, 2023/2024 – II pusmetis (4)



1 pav. Maskavimu paremto algoritmo veikimo schema¹



2 pav. Susiejimu paremto algoritmo veikimo schema¹

¹ <https://www.ceva-ip.com/ourblog/ai-audio-for-voice-enhancement-deep-into-the-deep-part-3/>

Moksliniai rezultatai, 2023/2024 – II pusmetis (5)

- Šiuo metu, mašininio mokymo metodų dėka, signalų apdorojimas yra nebūtinas.
- Taikant giliuoju mokymu grįstus sprendimus, galima išskirti dvi strategijas: su duomenų paruošimu, ir be jo.
- Buvo atlikta preliminari 33 publikacijų apžvalga, lyginami PESQ įverčio rezultatai.
- Maksimali įmanoma PESQ vertė yra 4.5. Publikacijose vidutinė PESQ vertė siekia:
 - Su apdorojimu **2.54**
 - Be apdorojimo **2.42**
- Aukščiausias PESQ rastas – 3.38. Darbe naudota požymių kombinacija iš MFCC, GFCC, AMS, RASTA-PLP.

Moksliniai rezultatai, 2023/2024 – II pusmetis (6)

- Žmogaus klausos sistemos požymių integravimas yra vienas iš galimų duomenų apdorojimo būdu treniravimo metu.
- Dažnai tokie požymiai naudojami nuostolių funkcijai apibrėžti.
- Pavyzdžiui, GAN modelyje diskriminatoriaus tinklas imituoja vertinimo metrikos elgesį, o generatorius generuoja tokį kalbos signalą, kurį diskriminatorius įvertintų gerai.
- Kitas būdas yra skaičiuoti tam tikrą dydį, atitinkantį žmogaus klausos savybes, ir aprašyti nuostolių funkciją taip, kad tas dydis būtų gerinamas.

Moksliniai rezultatai, 2023/2024 – II pusmetis (7)

- Pradėtas darbas VU DMSTI projekte – „Žmogaus suvokimo įkvėptas šnekos signalo gerinimas: gilaus mokymosi, triukšmo profiliavimo ir žmogaus klausos žinių integravimas“ (vadovė - dr. Gražina Korvel).
- Projekto metu gauti rezultatai ir išvados galės būti dalinai naudojami tyrimuose disertacijos metu.

Sekančio pusmečio darbo planas, 2024/2025 – I pusmetis

- Disertacijos tikslų ir uždavinių nustatymas.
- Citavimo rodiklio neturinčios publikacijos parengimas.
- Klasikinių bei mašininu mokymu grįstų kalbos gryninimo sprendimų palyginimas.

Klausimai