

Giliojo mokymo modeliai kalbos signalui gryninti

Parengė: Daniel Zakševski

Darbo vadovas: Doc. Dr. Gintautas Tamulevičius

Doktorantūros studijų metai: 2023-2027

Studijų metai: 2-ieji

Doktorantūros studijų ir mokslinių tyrimų plano bei jo vykdymo suvestinė

Studijų metai	Egzaminai	
	Planas	Ivykdyta
I (2023/2024)	2	2
II (2024/2025)	1	1
III (2025/2026)	1	
IV (2026/2027)		
Iš viso:	4	3

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

Ataskaitinio pusmečio darbo planas ir jo vykdymas, 2024/2025 – II pusmetis

Egzaminai, 2024/2025 – II pusmetis		
Planas	Ivykdyta	Būklė

Dalyvavimas konferencijose, 2024/2025 – II pusmetis		
Planas	Ivykdyta	Konferencijos tipas
	13-oji jaunųjų mokslininkų konferencija „Fizinių ir technologijos mokslų tarpdalykiniai tyrimai“	Nacionalinė

Publikacijos, 2024/2025 – I 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., 1.3. įvykdyti, 1.4. pradėtas vykdyti
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 mokymo 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 įvykdytas
2.1.	Tyrimo metodikos sudarymas:			
2.1.1.	Tyrimų tikslų ir uždavinių suformulavimas, tyrimo metodikos sudarymas	2024 m. II ketvirtis	2024 m. III ketvirtis	
2.2.	Teorinis tyrimas:			2.2.1. pradėtas vykdyti, 2.2.2. įvykdytas
2.2.1.	Klasikinių bei mašininio mokymo 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. pradėtas vykdyti
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	

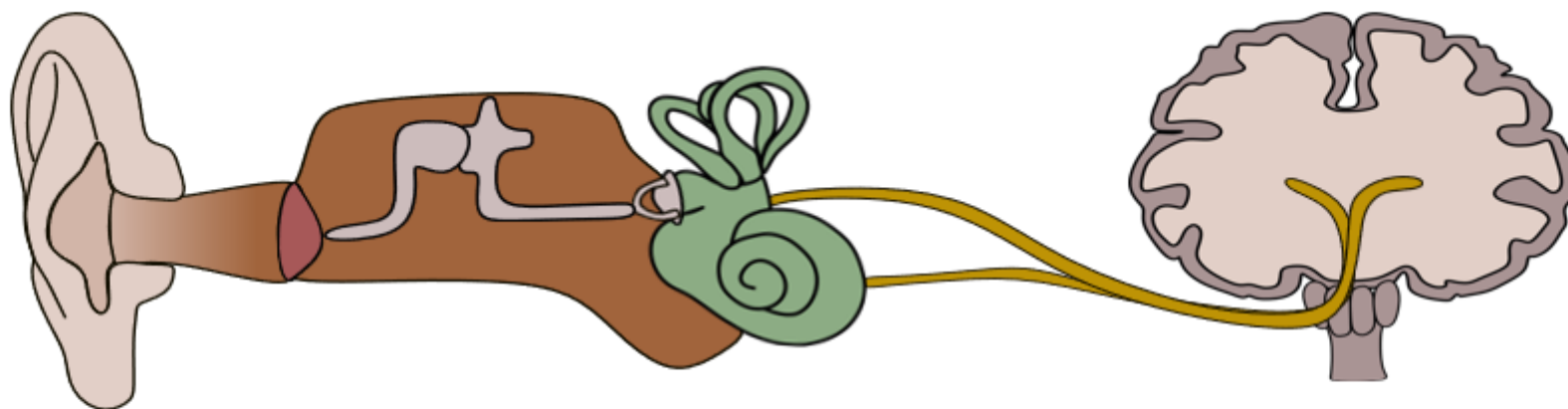
Tyrimo detalės

- **Tyrimo objektas:** šnekos signalas, balso signalo kokybė bei triukšmas.
- **Tyrimo tikslas:** pasiūlyti ir pritaikyti klausos modeliais grįstus giliojo mokymo tinklus kalbos signalui gryninti.
- **Tyrimo uždaviniai:**
 1. Įvertinti giliojo mokymo modelius naudojamus kalbos signalui gryninti.
 2. Apžvelgti ir įvertinti žmogiškosios klausos modelius.
 3. Integruoti klausos modelius, kitas žinias į kalbos signalo gryninimo modelį.
 4. Įvertinti pasiūlytųjų modelių modifikacijų efektyvumą, signalo gryninimo kokybę.

Moksliniai rezultatai, 2024/2025 – II pusmetis (1)

- Pagrindinis kalbos gryninimo (angl. speech enhancement) tikslas – padidinti akustinių kalbos signalų kokybę ir suprantamumą, pašalinant įvairius triukšmus bei jų sukeltus iškreipymus, atkuriant ar išryškinant spektrines kalbos signalo savybes.
- Geriausia šnekos gryninimo sistema yra žmogaus klausa. Savybių, leidžiančių žmogaus klausai efektyviai veikti, integravimas į skaitmeninius šnekos gryninimo algoritmus, potencialiai leistų toliau juos tobulinti.

Moksliniai rezultatai, 2024/2025 – II pusmetis (2)



	Išorinė ausis	Vidurinė ausis	Vidinė ausis	Smegenų kamienas ir smegenys
Funkcija:	Garso surinkimas	Vibracijų sustiprinimas	Dažnių analizė	Garso suvokimas ir interpretavimas

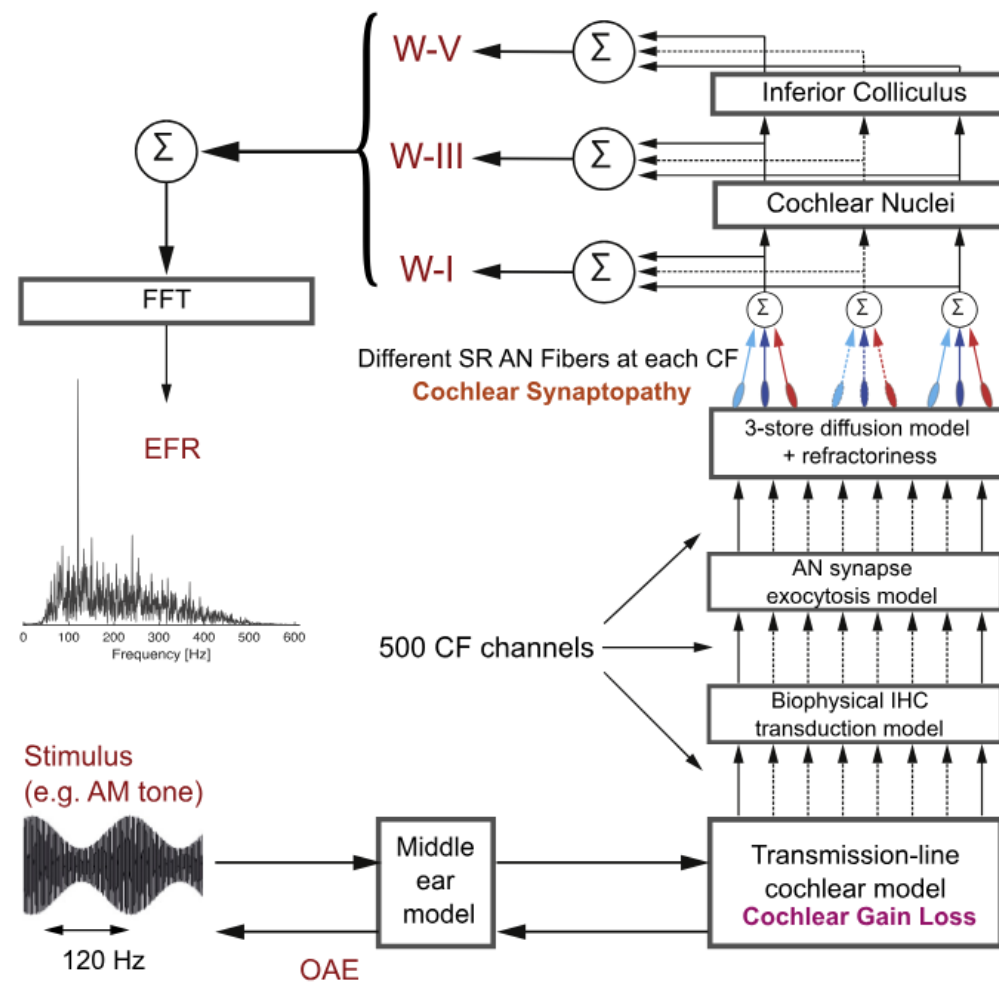
1 pav. Žmogaus klausos sistema, jos dalių funkcijos

Moksliniai rezultatai, 2024/2025 – II pusmetis (3)

- Žmogaus klausos sistemos požymių integravimas gali vykti įvairiose mašininio mokymo algoritmo dalyse:
 - Nuostolių funkcijose.
 - Tinklų architektūrose.
 - Požymių ištraukime.
- Klausos modelius galima grupuoti į: filtrų rinkinius (lygiagretus skaičiavimas) ir perdavimo linijos (nuoseklus skaičiavimas).

Moksliniai rezultatai, 2024/2025 – II pusmetis (4)

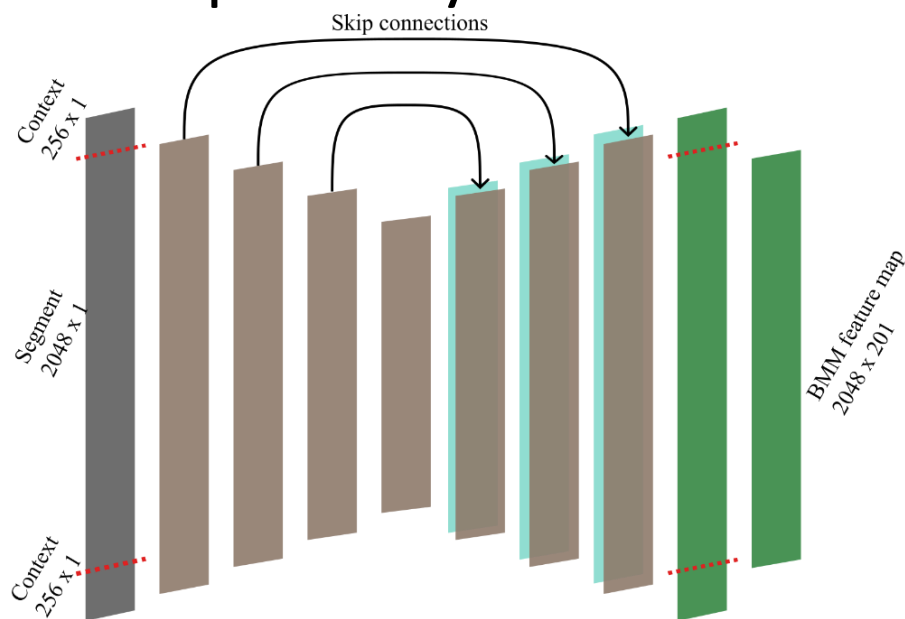
- Perdavimo linijos (angl. *transmission line*) modeliai tobulinami nuolatos. Vienas iš tokių klausos sistemos aprašymų sukurtas Verhulst et al. (2018) [1].
- Privalumas: itin detalus, fiziologiškai paremtas klausos modeliavimas.
- Trūkumas: sudėtingas, daug resursų reikalaujantis skaičiavimas.



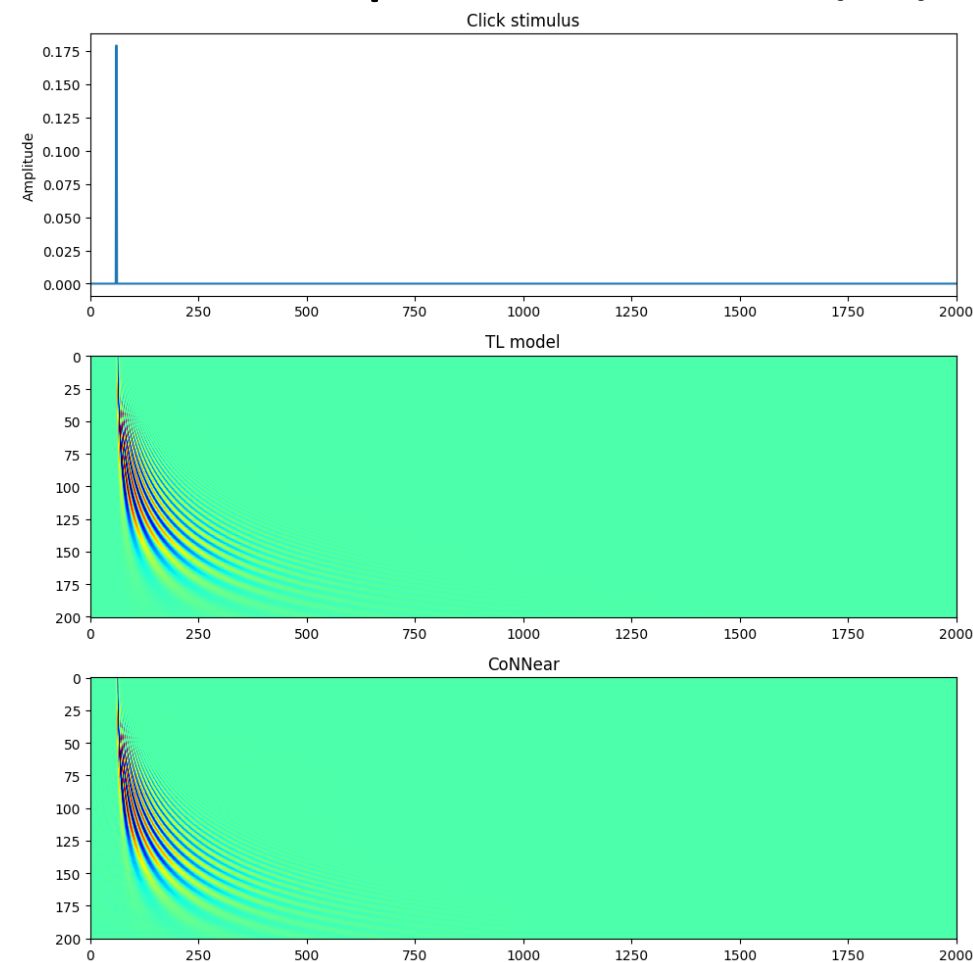
2 pav. Žmogaus periferinės klausos sistemos modelis [1]

Moksliniai rezultatai, 2024/2025 – II pusmetis (5)

- CoNNear – U-Net architektūros neuroninis tinklas. Išmokytas imituoti perdavimo linijos modelį. Stipriai pagreitina skaičiavimus, leidžia naudoti paskirstytuosius resursus.



3 pav. CoNNear neuroninio tinklo architektūra [2]



4 pav.

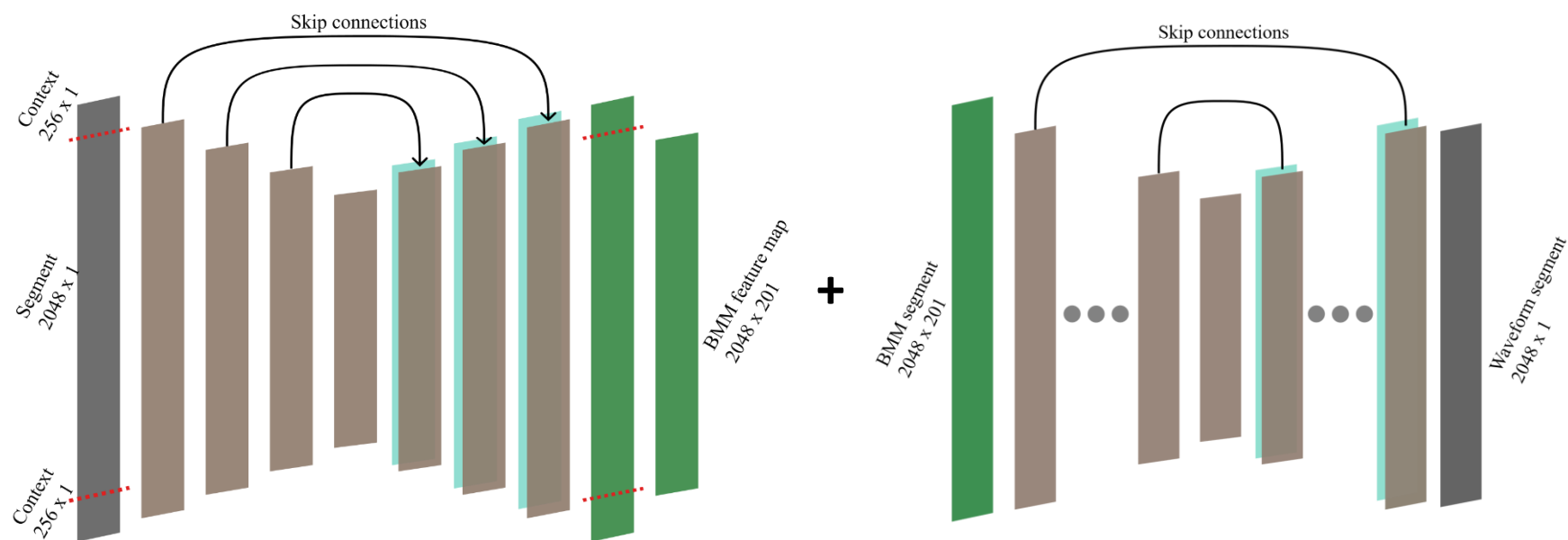
Viršutinis: Impulsinis signalas

Vidurinis: Perdavimo linijos modelio atsakas

Apatinis: CoNNear tinklo atsakas

Moksliniai rezultatai, 2024/2025 – II pusmetis (6)

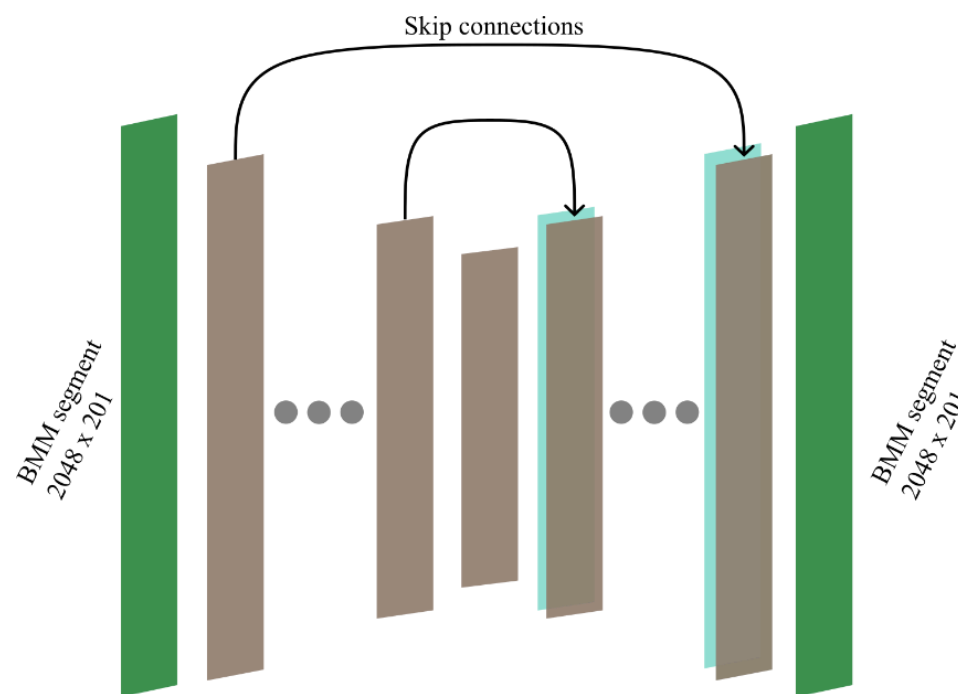
- CoNNear tinklu paremta triukšmo šalinimo sistemos pavyzdys yra CoSE – dar vienas U-Net architektūra paremtas neuroninis tinklas.
- Pažymėtina, kad AECNN atlieka ir triukšmo šalinimo, ir segmento rekonstravimo uždavinį.



5 pav. CoSE neuroninio tinklo architektūra. CoNNear + AECNN [3]

Moksliniai rezultatai, 2024/2025 – II pusmetis (7)

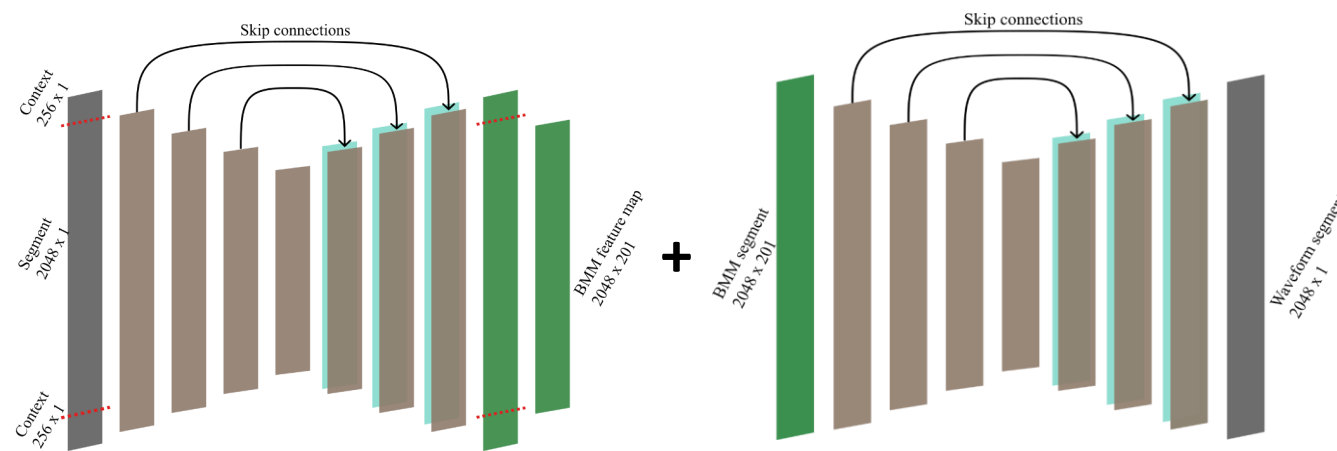
- Pavyko ne tik atkartoti autorių darbą, bet ir pasiūlyti patobulinimą: paskirti triukšmo šalinimo bei segmento rekonstrukcijos uždavinius atskiriems tinklams.



6 pav. deCoNNear neuroninio tinklo architektūra

Moksliniai rezultatai, 2024/2025 – II pusmetis (8)

- Ištirtas **CoNNear** -> **deCoNNear** veikimas, t. y. požymių išgavimas ir signalo sintezavimas be triukšmo šalinimo.



7 pav. CoNNear + deCoNNear testavimo sistemos architektūra

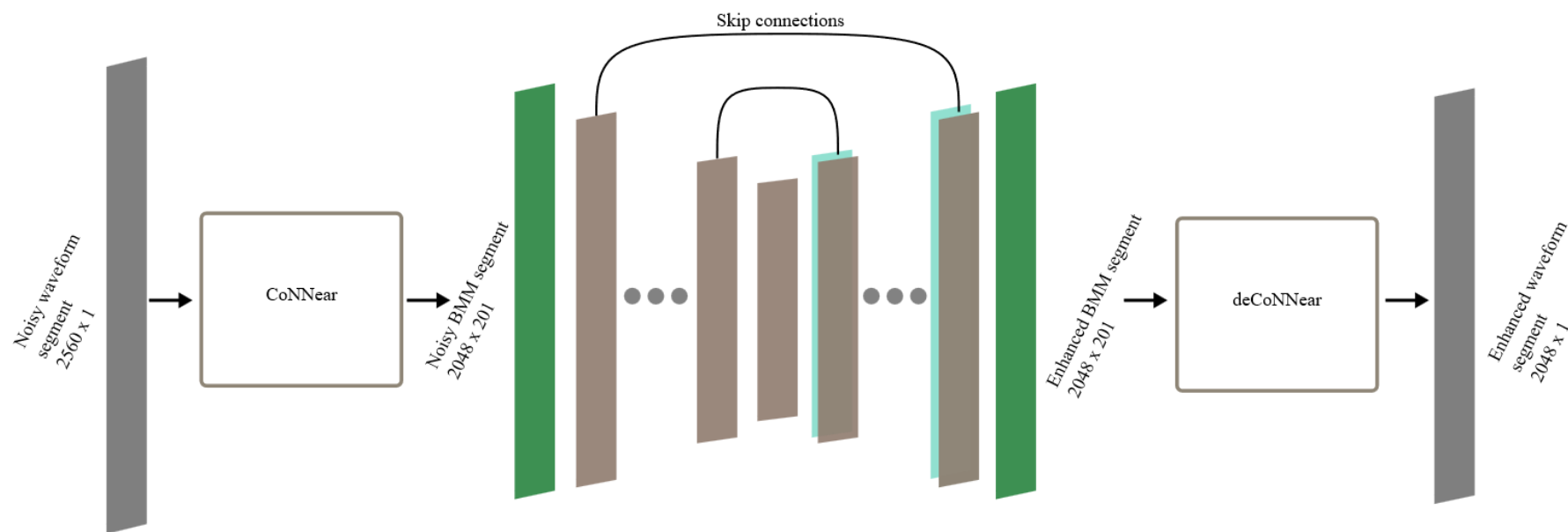
- Aukšti PESQ ir STOI rodo, kad originalus bei rekonstruotas įrašas beveik nesiskiria.

Metric	PESQ	STOI	SNR (dB)	SDR (dB)
Value	4.21	0.95	10.31	12.58

1 lentelė. CoNNear -> deCoNNear, be triukšmo šalinimo, rezultatai

Moksliniai rezultatai, 2024/2025 – II pusmetis (9)

- d-CoSE – pasiūlytoji sistema, kurioje triukšmas šalinamas požymių erdvėje (tarp BMM), o išvalytus požymius rekonstruoja deCoNNear.



8 pav. Pasiūlytoji d-CoSE šnekos gryninimo sistema. Sudaryta iš CoNNear, AECNN, deCoNNear

Moksliniai rezultatai, 2024/2025 – II pusmetis (10)

Model	Loss function	SNR, dB	SSNR, dB	PESQ	STOI
Noisy	-	8.39	-0.50	1.94	0.76
CoSE	L1_waveform	-0.06	-0.90	<u>2.33</u>	0.93
CoSE	L1_STFT	-1.97	-3.36	1.98	0.91
d-CoSE	L1_waveform	0.18	-1.04	2.26	0.92
d-CoSE	L1_STFT	-19.98	-25.70	1.66	0.92
d-CoSE	L1_BMM	-9.61	-16.61	1.69	0.91
d-CoSE ^{FT}	waveform + BMM	-0.16	-0.72	2.18	0.92
d-CoSE ^{FT}	waveform + STFT + BMM	0.59	-1.34	<u>2.30</u>	0.93

2 lentelė. CoSE ir d-CoSE šnekos gryninimo sistemų palyginimas esant skirtingoms nuostolių funkcijoms.

FT reiškia „fine-tuned“.

Parodytos vidutinė įverčio reikšmė visam testavimo rinkiniui.

Moksliniai rezultatai, 2024/2025 – II pusmetis (11)

- Pristatyta atskirta d-CoSE sistema, skirta biofizikiniu principu pagrįstam šnekos gryninimui, ir palyginta su monolitine CoSE sistema.
- Modulinis metodas parodė veiksmingą, kokybišką signalo rekonstrukciją naudojant CoNNear -> deCoNNear.
- d-CoSE, apmokymui naudojant hibridinę nuostolių funkciją, pasiekė 2,3 PESQ įvertį, panašų rezultatą, kaip originaliosios monolitinės sistemos 2,33.
- Tai rodo, kad uždavinių atsiejimas turi privalumus, tokius kaip moduliarumas, modelio paaiškinamumas ir pažangių mokymo metodų taikymas – esant nežymiam signalo kokybės rodiklių kritimui.

Sekančio pusmečio darbo planas, 2025/2026 – I pusmetis

- Pranešimas tarptautinėje konferencijoje 6th International Conference on Communications, Information, Electronic and Energy Systems (CIEES 2025).
- Žmogaus klausos sistema pagrįstų giliojo mokymo empirinio tyrimo tęsimas.
- Išlaikyti dalyko „Kalbos signalo apdorojimas“ egzaminą.

Klausimai

Šaltiniai

- 1. S. Verhulst, A. Altoè, and V. Vasilkov, “Computational modeling of the human auditory periphery: Auditory-nerve responses, evoked potentials and hearing loss,” *Hear Res*, vol. 360, pp. 55–75, Mar. 2018, doi: 10.1016/J.HEARES.2017.12.018.
- 2. D. Baby, A. Van Den Broucke, and S. Verhulst, “A convolutional neural-network model of human cochlear mechanics and filter tuning for real-time applications,” *Nature Machine Intelligence 2021 3:2*, vol. 3, no. 2, pp. 134–143, Feb. 2021, doi: 10.1038/s42256-020-00286-8.
- 3. C. Wen and S. Verhulst, “Biophysically-inspired single-channel speech enhancement in the time domain,” *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, vol. 2023-August, pp. 775–779, 2023, doi: 10.21437/INTERSPEECH.2023-936.