**

**VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS**

**FUNDAMENTINIŲ MOKSLŲ FAKULTETAS**

**GRAFINIŲ SISTEMŲ KATEDRA**

Gleb Čirkov, Maksim Igošev, Osmundas Marcinkonis

MKDfs-18

**Skaitmeninių Signalų Apdorojimas: balso komandų atpažinimas naudojant Dynamic Time Warping metodą**

Praktinės užduoties aprašas

Turinys

[Įvadas 3](#_Toc37535061)

[Užduotis ir sprendimas 3](#_Toc37535062)

[Užduoties pasirinkimas 3](#_Toc37535063)

[Sprendimo panaudojimas 3](#_Toc37535064)

[Planuojami Rezultatai 4](#_Toc37535065)

[Sprendimo kūrimo priemonės 4](#_Toc37535066)

[Reikalavimai 4](#_Toc37535067)

[Funkcionalumas 4](#_Toc37535068)

[Naudotojo sąsaja 5](#_Toc37535069)

[Naudojama metodika 7](#_Toc37535070)

[Dynamic Time Warping Algoritmas 7](#_Toc37535071)

[Algoritmo veikimas 8](#_Toc37535072)

[Šaltiniai 11](#_Toc37535073)

# Įvadas

## Užduotis ir sprendimas

Pasirinkta skaitmeninių signalų apdorojimo modulio praktinė užduotis - sukurti programinę įrangą, kuri būtų pritaikyta įrašyti ir atpažinti vartotojo balso komandas. Taip pat siekiama optimizuoti programinę įranga taip, kad ji naudotų kuo mažiau atminties, tuo pačiu nenukenčiant jos pagrindiniam funkcionavimui. Siekiama, kad galutinis produktas galėtų būti talpinamas į mažesnio pajėgumo mobilius prietaisus (pvz.: Arduino, Nanode, Raspberry Pi). Sprendimo įgyvendinimui nuspręsta pasitelkti Dynamic Time Warping metodą[1][2].

## Užduoties pasirinkimas

Išmaniesiems telefonams tapus daugelio žmonių kasdienybe, stipriai populiarėja virtualių asistentų (Google Assistant, Siri, Amazon Home, Cortana) naudojimas visuomenėje. Stebint šių technologijų populiarumo augimą, darosi įdomu asmeniškai suprasti kaip veikia analoginio žmogaus balso skaitymas, mašininis analizavimas bei interpretavimas. Supratus šio proceso pagrindinius etapus, tikėtina, kad taps suprantamos ir šio proceso pagrindinės problemos (pvz.: nevisiškas balso atpažinimo tikslumas dėl žmogiškųjų faktorių).

Taip pat, žvelgiant į užduotį iš multimedijos pusės, kuriant balso atpažinimo sistemą gali paaiškėti, kodėl kompiuterinių žaidimų industrija retai arba nesėkmingai naudoja šią technologiją.

## Sprendimo panaudojimas

Sukurtas sprendimas, dėl savo gebėjimo prisitaikyti prie vartotojo balso naudojant Dynamic Time Warping metodą, turi ganėtinai platų naudojimo spektrą. Įrašius programinę įrangą išmaniajame telefone, galutinis produktas gali atitikti virtualaus asistento darbą arba, sukūrus atitinkamą sąsają, gali vykdyti vartotojo pasirinktas komandas darbui su technine įranga (pvz.: įjungti/išjungti apšvietimą tam tikroje vietoje). Sprendimas taip pat gali būti taikomas neįgaliųjų vartotojų alternatyviai įvesčiai. Taip pat, kaip anksčiau minėta, tai gali būti naudojama multimedijoje (kompiuteriniuose žaidimuose, meno instaliacijose ir t.t.).

## Planuojami Rezultatai

Tikimasi, kad sukurtas produktas gebės veikti įvairiose aplinkose (iOS, Windows, Linux ir t.t.) bei prietaisuose (kompiuteriuose, telefonuose, mikrokontroleriuose). Gebės įrašyti ir sparčiai atpažinti vartotojo balso komandas, bei balso komandai priskirtus veiksmus.

## Sprendimo kūrimo priemonės

Sprendimo kūrimui pasitelkta MATLAB programavimo aplinka dėl savo universalumo; bei Dynamic Time Warping atpažinimo metodas dėl greito veikimo, nenaudojant daug kompiuterinių resursų.

# Reikalavimai

Ši programa gali būti realizuota įvairiomis formomis, tačiau šiuo metu naudojamas tinklalapis, nes taip programinė įranga tampa plačiai ir greitai prieinama. Šiuo metu šiai programinei įrangai planuojami tokie reikalavimai:

* Prieiga prie interneto (jei programinė įranga nėra atsisiųsta).
* Prieiga prie mikrofono (audio įvesties).
* Prieiga prie garsiakalbio (audio išvesties).
* Vartotojas turi gebėti skaityti ir raiškiai kalbėti.

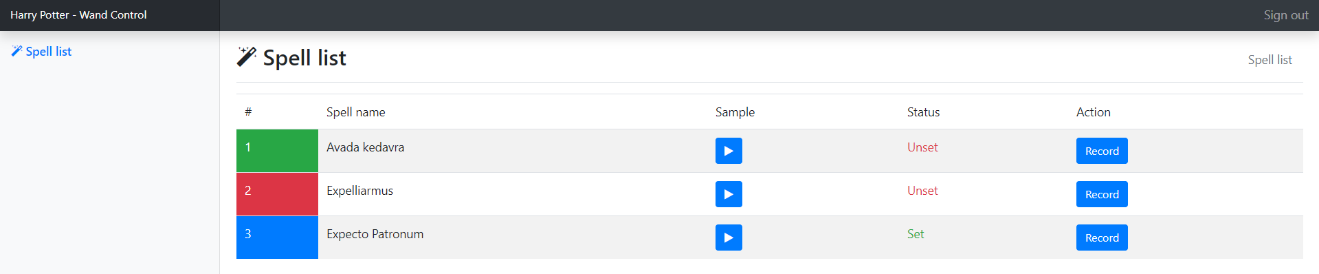
# Funkcionalumas

Kuriama programinė įranga turi šiais pagrindines funkcijas:

* Vartotojo komandų balso išsaugojimas.
* Vartotojo pasakytos komandos atpažinimas iš išsaugotos bibliotekos.
  + Gauto audio failo bei bibliotekos įrašų vertimas į skaičiavimo matricas.
  + Audio failų (gauto ir bibliotekos) lyginimas Dynamic Time Warping metodu.
  + Panašiausio įrašo nustatymas.
* Veiksmo, susieto su sutapusia komanda, vykdymas. (Pvz.: užsidega tam tikros spalvos lemputė).

# Naudotojo sąsaja

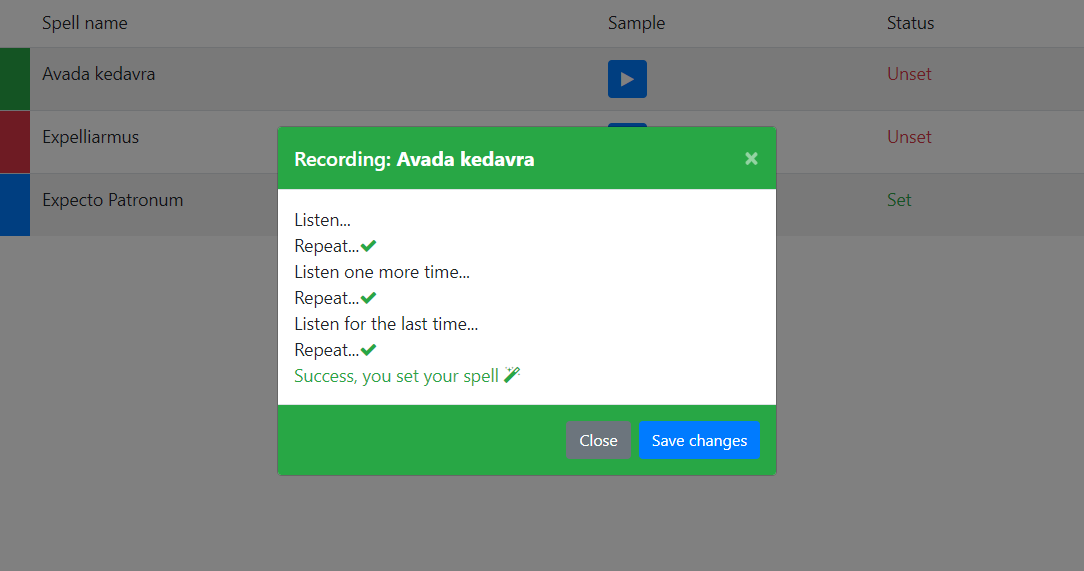
Šiai programinės įrangos prototipo grafinei vartotojo sąsajai buvo parinkta *Hario Poterio* tematika ir balso komandos sutapatinamos su *burtažodžiais*. Taip pat yra naudojama tarptautinė, anglų kalba.

 Vartotojo sąsaja (*pav. 1*) grafiškai perteikia programos funkcijas, rodo burtažodžių būsena bei nurodo vartotojui veiksmų seką, kuri reikalinga programos įgalinimui.

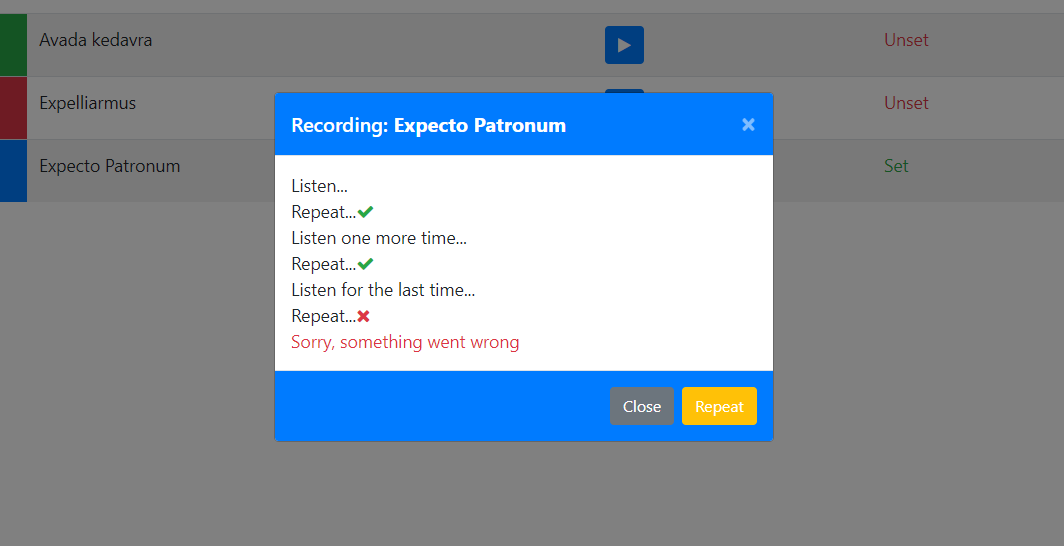
pav. : vartotojo sąsajos prototipas.

Kadangi programinė įranga nėra sudėtinga, grafinėje sąsajoje yra tik du mygtukai: „Sample“ ir „Record“. „Sample“mygtukas, atvaizduota paleidimo simboliu – baltas trikampis mėlynam fone - paleidžia garso įrašą per vartotojo garsiakalbį, kaip pavyzdį, kaip reikėtų tarti burtažodį. „Record“ mygtukas atidaro įrašymo langą, kuriame vartotojas vedamas per burtažodžio įrašymo į biblioteką procesą.

Įrašymo lange vartotojui tris kartus yra paleidžiamas pavyzdinis burtažodžio įrašas, po kiekvieno karto prašant pasakyti burtažodį į mikrofoną. Tai atlikus taisyklingai, pateikiama žinutė, kad procesas įvyko sėkmingai (*pav. 2*). Priešingu atveju – nurodoma, kad įvyko klaida ir prašoma kartoti operaciją (*pav. 3*). Sistema rodo klaidą tada, kai vartotojas pateikia burtažodžius, kurie yra nepanašūs tarpusavyje.



pav. : sėkmingai pateiktas burtažodis.



pav. : trečią kartą vartotojo pateiktas burtažodis nepanašus į pirmus du, prašoma pakartoti įvestį.

# Naudojama metodika

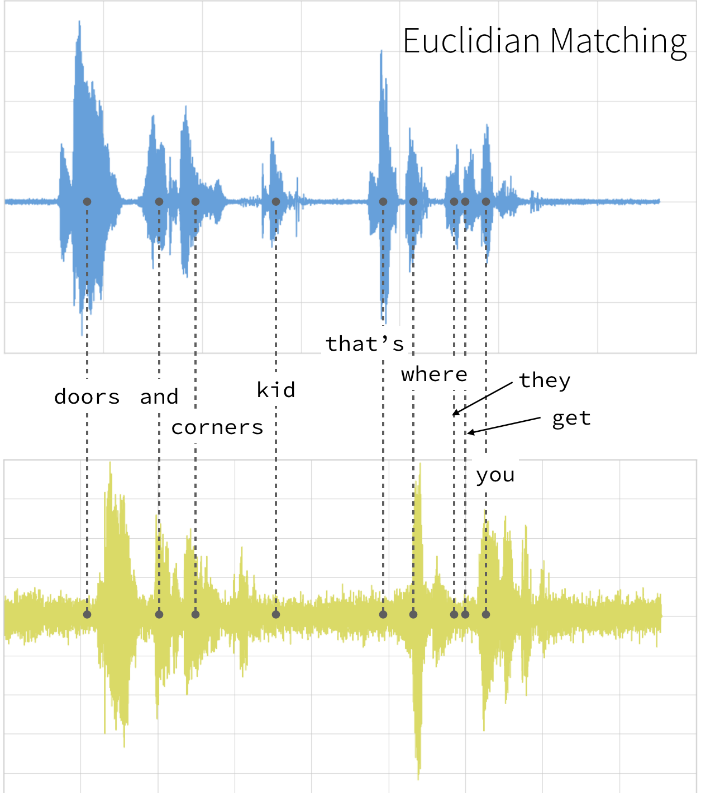
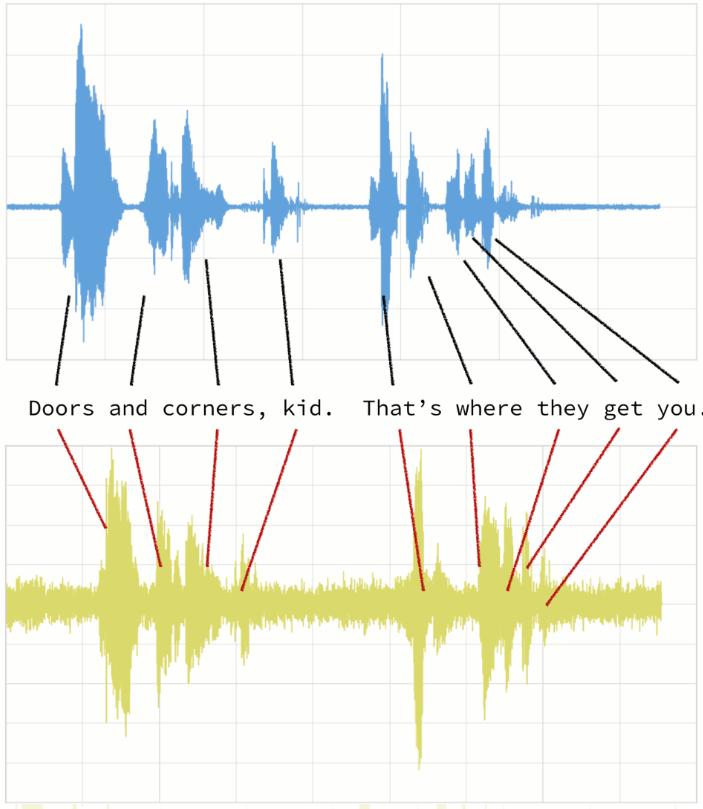
Šiame skyriuje aptariamas pagrindinis garso įrašo lyginimo algoritmas – Dynamic Time Warping.

## Dynamic Time Warping Algoritmas

Dynamic Time Warping (toliau - DTW) yra algoritmas, skirtas įvertinti koreliaciją tarp dviejų duomenų rinkinių, kurių laiko parametras yra nevienodo dydžio. Šis algoritmas gali būti pritaikytas duomenų analizei daugelyje sričių - lyginti finansinei statistikai tarp mėnesių, kurių dienų skaičius nevienodas (pvz. balandis - 30 dienų, gegužė - 31 diena); žingsniamačiuose - nueitų žingsnių kiekiui skaičiuoti, kai naudotojo ėjimo greitis nėra stabilus ir pan. [3]

Šiame darbe DTW bus taikomas kalbos atpažinimui - bus ieškoma koreliacijos tarp nevienodos trukmės įrašų. Vartotojas pateiks tris balso įrašus, kurių turinys privalės būti vienas ir tas pats burtažodis. Šie įrašai, patvirtinus, kad yra užtektinai panašūs, bus laikomi etalonais - sistemai suprantamais burtažodžiais. Toliau dirbant su programa, didžiausią koreliaciją turintis etalonas apspręs, kokį žodį sistemai pateikė vartotojas. Galios sąlyga, kad turi būti pasiekta minimali koreliacijos reikšmė. Nepasiekus šios reikšmės, sistema leis vartotojui žinoti, kad jo sakomas žodis (burtažodis) nėra tinkamas. Pasakius tinkamą burtažodį, užsidegs atitinkamos spalvos lemputė, priskirta konkrečiam burtažodžiui.

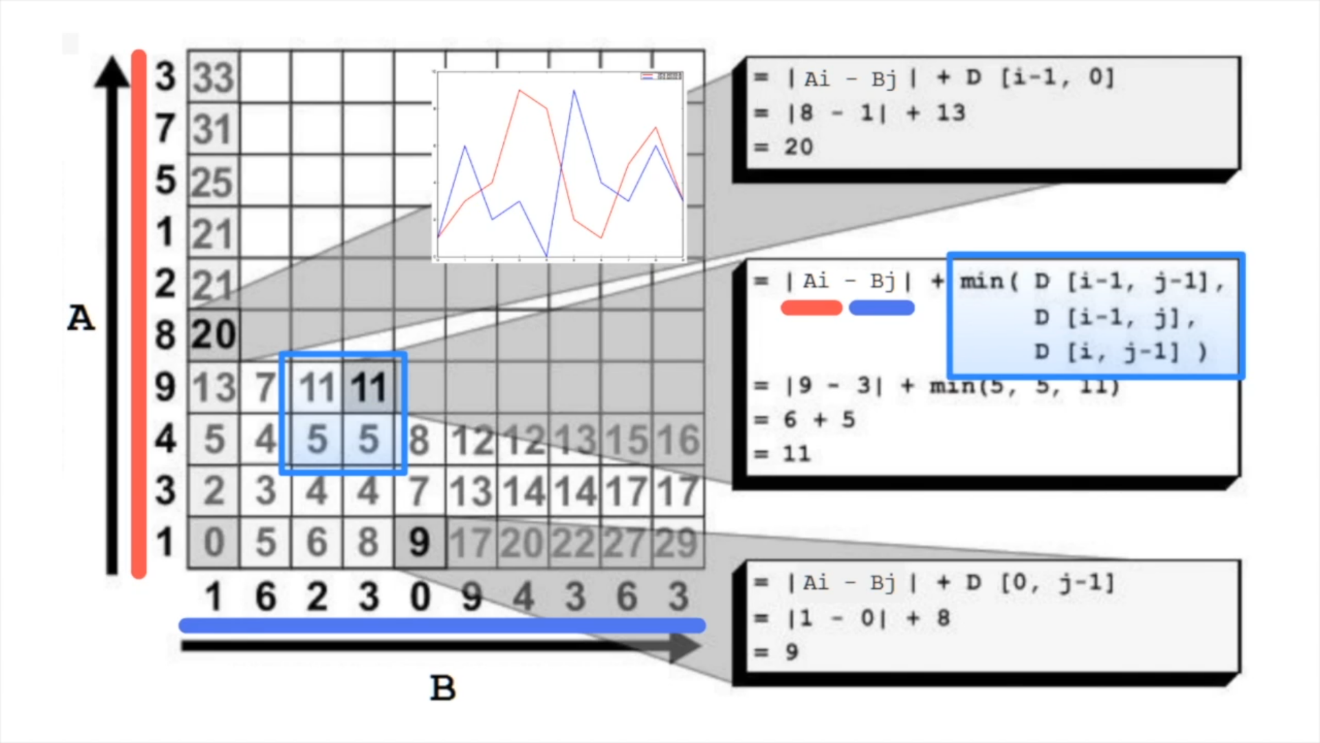
DTW metodas pasirinktas, nes išsprendžia nevienodo kalbos greičio problemą. Ieškant koreliacijos nuosekliai, kalbos atpažinimas nėra patikimas. DTW yra veiksmingas sprendimas, nes leidžia manipuliuoti laiku (*pav. 4* – nuosekli paieška, *pav. 5* – DTW)

****

pav. : nuosekli koreliacijos paieška.

pav. : Dynamic Time Warping.

## Algoritmo veikimas

**** Iš vartotojo pateikto įrašo ir etalonų sudaromos skaičiavimų matricos (pagal *pav. 6* pateiktas formules). Žemiau esančiame pavyzdyje nurodyta, kaip gauti skaičiai 20, 11 ir 9.

pav. : matricos gavimas pagal pateiktas formules.

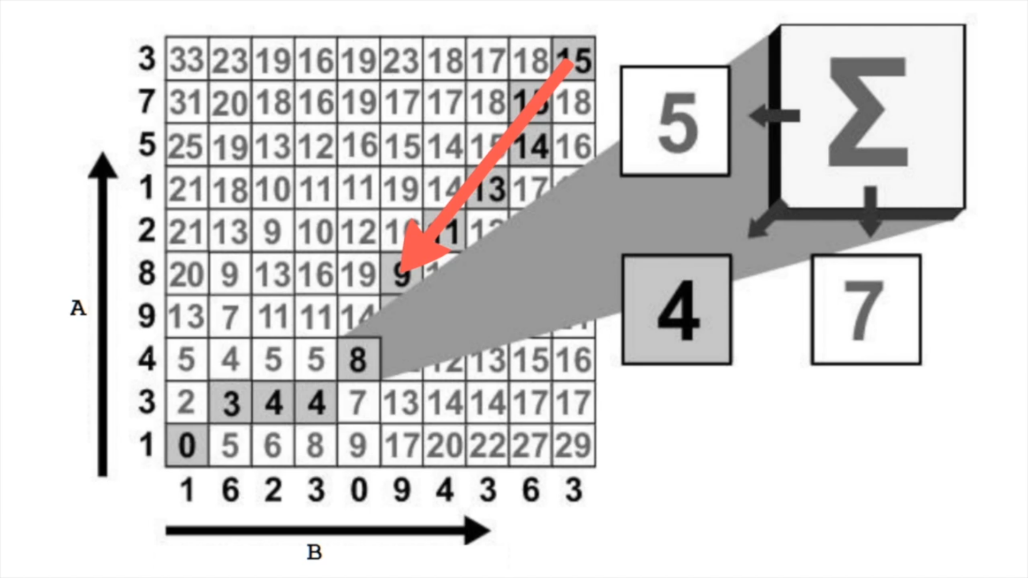
Skaičiuojami atstumai tarp dviejų įrašų amplitudžių. Paimamos atitinkamos reikšmės iš įrašo A ir įrašo B. Jos viena iš kitos atimamos, tada taikomas modulis. Galiausiai, pridedama mažiausia reikšmė iš ankstesnių skaičiavimų. Reikšmė gali būti pasirinkta iš anksčiau apskaičiuotų reikšmių esančių šalia: iš kairės, apačioje, arba įstrižai kairėn-žemyn.

Pavyzdyje - skaičius 20 gautas iš 8 atimant 1 ir pridedant 13, vienintelę tinkamą anksčiau apskaičiuotą reikšmę (anksčiau gautų reikšmių kairėje ir įstrižai - nėra). Gauta lygybė: |8-1|+13=20.

11 gautas iš 9 atimant 3 ir pridedant mažiausią tinkamą ankstesnį rezultatą (galimi skaičiai - 11, 5 ir 5; pasirinktas mažiausias, t.y. 5). Gauta lygybė: |9-3|+5=11.

9 gaunamas analogišku būdu. Iš 1 atimamas nulis, taikomas modulis. Vienintelė tinkama pridėti reikšmė yra kairėje ir tai yra 8. Apačioje ir įstrižai tinkamų reikšmių nėra. Gauta lygybė: |1-0|+8=9.

Tokiu principu, gaunama visa matrica. Kai gaunama pilna matrica, ieškoma trumpiausio kelio nuo viršutinio dešiniojo matricos taško iki apatinio kairiojo (*pav. 7*). Pradedama nuo viršutinio dešiniojo kampo, pasirenkama mažiausia reikšmė iš 3 šalia esančių. Taip keliaujama iki apatinio kairiojo kampo.

****

pav. : trumpiausio kelio paieška.

Viską atlikus, surandamas mažiausias įmanomas atstumas tarp dviejų įrašų amplitudžių. Mažiausias atstumas atitinka didžiausią koreliaciją. Pagal tai nustatoma, kuris įrašas panašiausias. Jei panašumas užtektinai aukštas, sistema laiko vartotojo pasakytą žodį atpažintu ir vykdomas atsakas - užsidega tam tikros spalvos lemputė, priklausomai nuo to, kuris žodis buvo pasakytas vartotojo.

# Šaltiniai

[1] Sakoe, H., Chiba, S. (1978). *Dynamic programming algorithm optimization for spoken word recognition.* IEEE Trans. Acoustic Speech and Signal

[2] Souza, C.F.S., Pantoja, C.E.P, Souza, F.C.M. Verificacao de assinaturas offline utilizando Dynamic Time Warping. Proceedings of IX Brazilian Congress on Neural Networks, v1 pp. 25-28. 2009.

[3] [**https://databricks.com/blog/2019/04/30/understanding-dynamic-time-warping.html**](https://databricks.com/blog/2019/04/30/understanding-dynamic-time-warping.html)