

Home Page

Title Page

◀ ▶

◀ ▶

Page 1 of 19

Go Back

Full Screen

Close

Quit

0.1. Fiksuoto teksto ir diktoriaus atpažinimas

Paanalizuosime detaliau fiksuoto teksto ir diktoriaus atpažinimą. Remiantis šiame skyrelyje pateikta informacija galėsite atlikti balso atpažinimo temai skirtą praktinę užduotį. Eksperimentams naudosi-
me **ELRA-S0050 Russian speech database (STC)** duomenų bazės poaibį. Sieksime išmokti atpažinti
dvi rusų kalbos frazes "Nie ishcite stenografistku jiesli u vas est' transkraiber i zielanijie rabotat" ir
"Cifrovojie ustroistvo sumoatchistki lutcij pomoscnik v liubych zatrudnienijach" ir nustatyti ar jos yra
sakomos to paties diktoriaus ar ne. Kad išspręsti šį uždavinį, mums reikės atlikti tokius žingsnius:

- Nuskaityti balso duomenis;
- Išskirti balso požymius;
- Sukonstruoti dviejų balso pavyzdžių požymių palyginimo metriką;
- Realizuoti etaloninių balso pavyzdžių požymių saugyklą;
- Išsaugoti palyginimo rezultatus unifikuotu formatu, kad būtų galimybė grafiškai vizualizuoti algoritmo kokybę DET kreive.

Lengviausiai atliekamas paskutinis žingsnis - jei jau esate realizavę veidų, rainelių, pirštų atspaudų, parašų palyginimo užduotį, tai ši dalis nieko nesiskiria nuo ankstesnių palyginimo rezultatų, kuriuos

išsaugodavote **.roc** formatu.

Kad nuskaityti balso duomenis, galite susirasti patys tinkamas skaityti **.wav* formato failams arba pasinaudoti **WaveReader** java kodu. Teisingai nuskaite balso duomenis, turėtumėte gauti **1** paveikslėlyje pavaizduotą balso duomenų grafiką. Iliustracijai naudoto *010101.wav* ir kitus balso pavyzdžių duomenis rasite **čia**.

Turint balso skaitmeninius duomenis, galima išskirti juose požymius. Požymiams išskirti mes naudosome balso duomenų **spektrogramą**. Spektrogramos apskaičiavimui galite pasirinkti jums tinkančią **biblioteką** arba pasinaudoti **AsimetEkspo** java kodu. **AsimetEkspo.main** metode rasite spektrogramos apskaičiavimo ir jos iliustracijos pavyzdį. **2** pav. iliustruoja *AsimetEkspo* programa gautą *010101.wav* duomenų spektrogramą.

Tame pačiame *AsimetEkspo.java* faile pateiktas kodas, kuris remiantis spektrograma, išskiria joje formantes. Tai gaunama tokiu būdu:

- atliekamas spektrogramos vidurkinimas dažnio kryptimi pasirinktu masteliu;
- atliekamas pakartotinis vidurkinimas kitu masteliu;
- apskaičiuojamas dviejų skirtingais masteliais vidurkintų spektrogramų skirtumas.

Home Page

Title Page

◀ ▶

◀ ▶

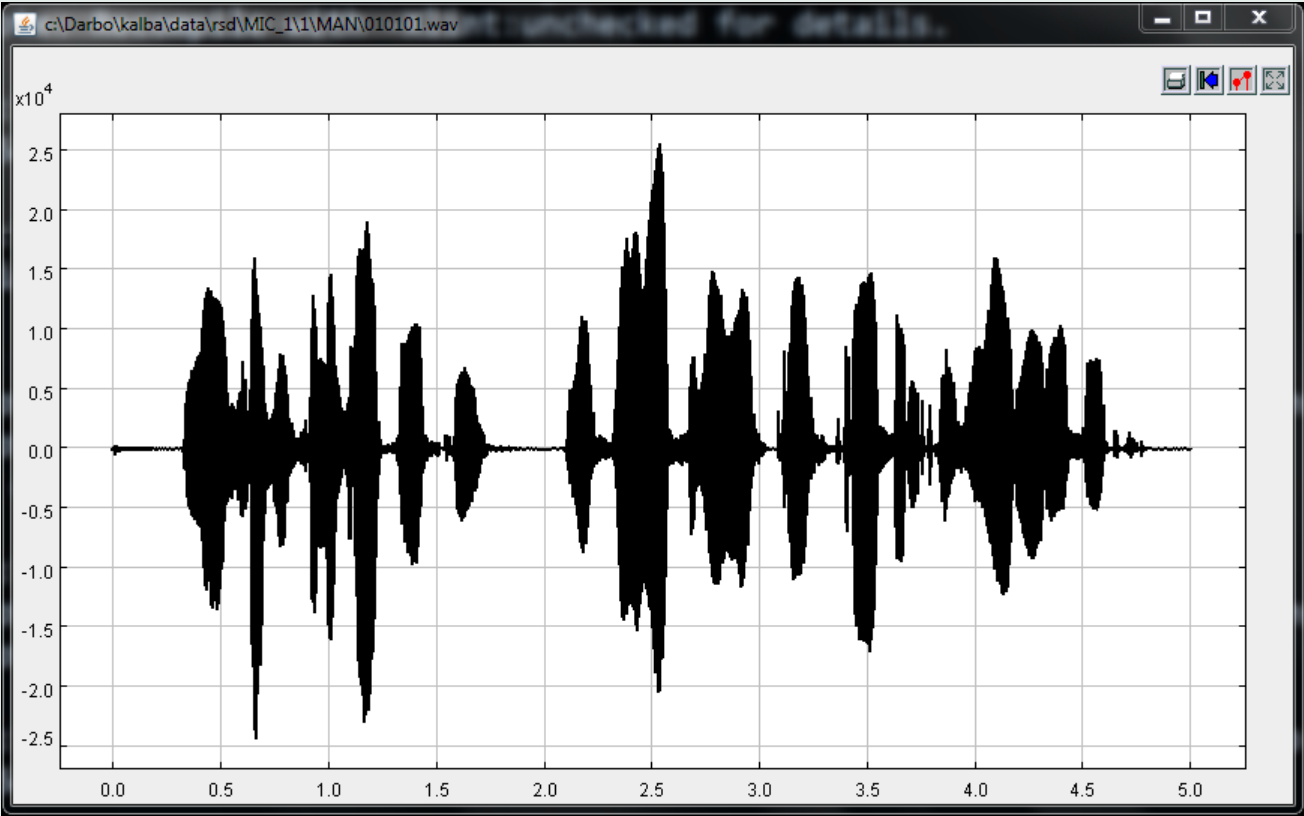
Page 3 of 19

Go Back

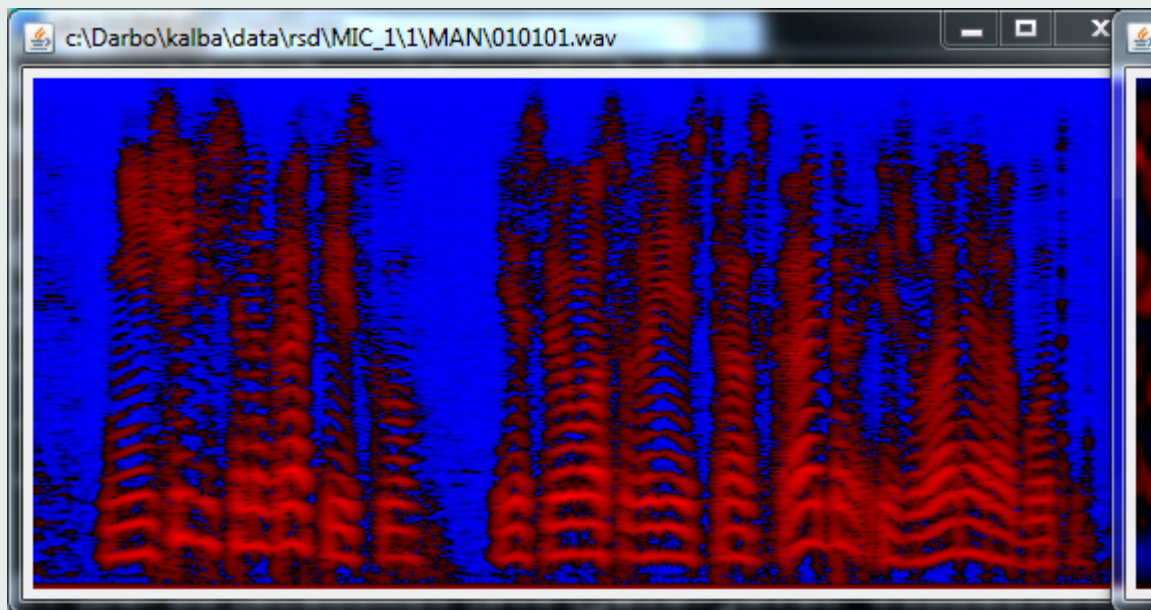
Full Screen

Close

Quit

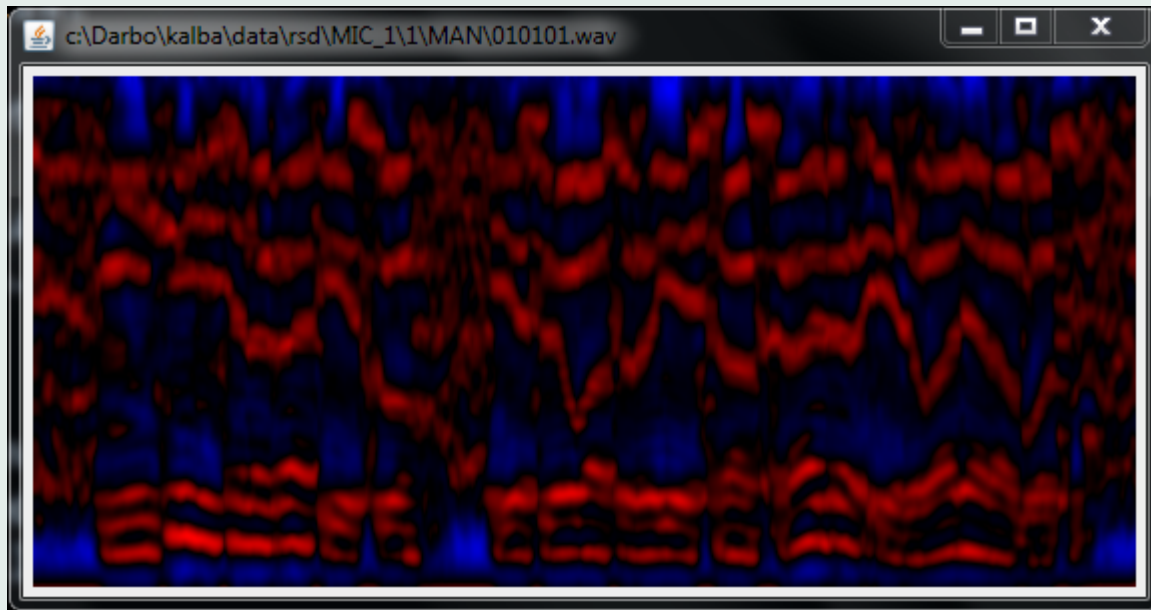


1 pav.: 010101.wav balso duomenų grafikas



2 pav.: 010101.wav balso duomenų pavyzdžio spektrograma

Tokiu būdu gautame vaizde išryškėja formantės. 3 pav. iliustruoja 010101.wav spektrogramoje išryškintas formantes. Formantės atitinka raudoniausias vaizdo vietas. Iš grafiko matyti, kad labiausiai informatyvi antroji nuo apačios formantė, kuri svyruoja apie 3000 Hz. Tikriausiai neatsitiktinai žmogaus klausa labiausiai jautri apie 3000 Hz. dažniams. Abscisių ašis ir spektrogramoje ir formančių vaizdelyje turi laiko dimensiją, ordinačių - dažnius. Kairysis apatinis taškas vaizduoja pradinį mo-



3 pav.: Spektrogramoje išskirtos formantės

mentą ir nulinį dažnį. Vaizdelio taškų spalva ir intensyvumas iliustruoja spektro arba formantės intensyvumą duotu laiko momentu ir duotame dažnyje, kurių reikšmės nusakomos taškelio vaizde pozicijos koordinatėmis.

Turint formantes, galimą pereiti prie požymių skaičiavimo. Vienas būdų išskirti požymiams bū-
tų aprašyti kreivėmis formantes. Pavyzdžiui per formančių centrines dalis nubrėžti Bezjė kreives ir

*Home Page**Title Page*

◀◀

▶▶

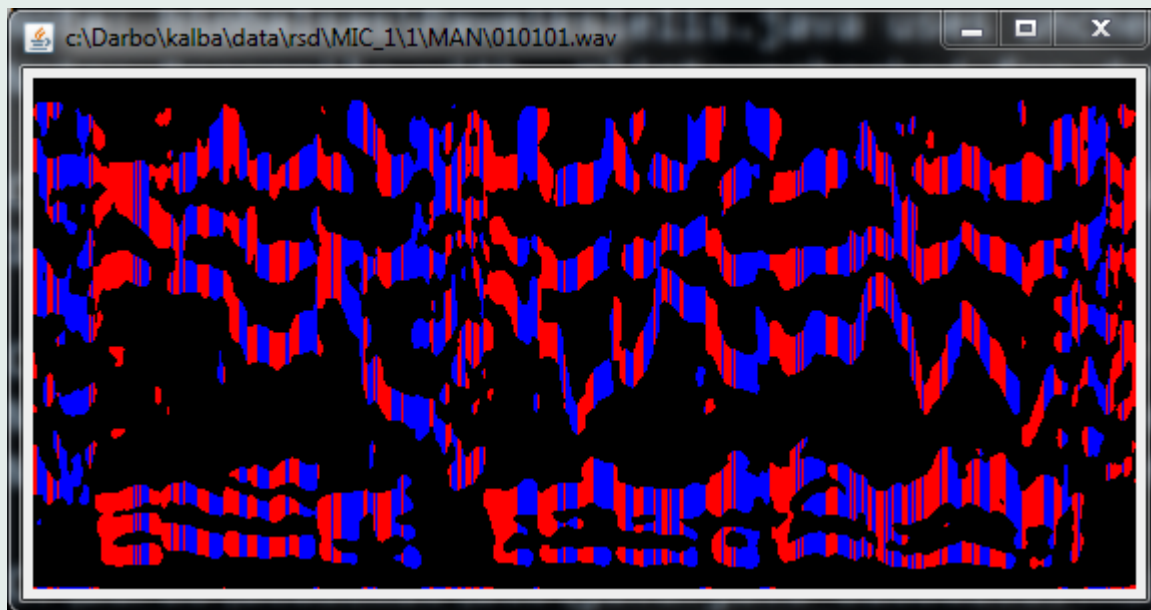
◀

▶

*Page 6 of 19**Go Back**Full Screen**Close**Quit*

požymiais laikyti Bezjė kreives aprašančius parametrus. Tačiau tai būtų gana paini procedūra; mes pasielgsime paprasčiau - tiesiog binarizuosime formančių vaizdelį. 4 paveikslėlis iliustruoja binarizuos formantes. Papildomai binarizuotos formantės spalvinamos įvertinant žodžių pradžias ir pabaigas. Pradžios spalvinamos raudonai, pabaigos - mėlynai. Pradžia ir pabaiga vertinama pagal eksponentinio filtravimo pirmyn ir atgal gaunamų reikšmių santykinę energiją: jei duotame laiko taške didesnė filtravimo pirmyn energija, laikoma, kad tai žodžio pabaigos dalis. Vizualiai matyti, kad laiko segmentuose, kurie atitinka tariamo balso pauzes (tylos momentus), formantės pasislinkusios aukščiau. Siūloma savarankiškai pagal pirmųjų formančių pasislinkimą įvertinti balso pauzių momentus ir vizualizuojant naudoti papildomas spalvas, kad išskirti balso pauzių laiko intervalus.

6 paveikslėlis iliustruoja "kvadra" požymius, kurie gaunami apjungiant spektrogramos taškelio du bitus informacijos (čia žemiausias dažnis atitinka paveikslėlio viršutinę eilutę). Pirmas bitas pažymi ar skirtumas vidurkintų skirtingais masteliais spektrogramos einančio per taškelį stulpelio yra teigiamas. Antras bitas pažymi ar spektrogramos reikšmės taškelyje su reikšme paimta paslinkus per tris taškelius laiko kryptimi yra teigiamas. Šie du bitai informacijos išreiškiami keturiomis spalvomis: 11 - raudona, 01 - mėlyna, 01 - geltona ir 00 - ciano.



4 pav.: 010101.wav balso duomenų binariniai požymiai

0.2. Požymių palyginima

Požymių palyginimui naudosime panašią į parašų požymių palyginimo techniką. Pirma mums reikia įvertinti dviejų spektrogramų stulpelių panašumas. Tai atliekama suskaičiavus stulpeliuose sutampančių bitų kiekį. Kadangi pavyzdžiuose spektrogramos stulpelio aukštis 256 taškai ir vienas taškelis aprašomas dviem bitais informacijos, viso turime sulginti $256 \times 2 = 512$ bitų. Kad tai greitai padaryti,

Home Page

Title Page



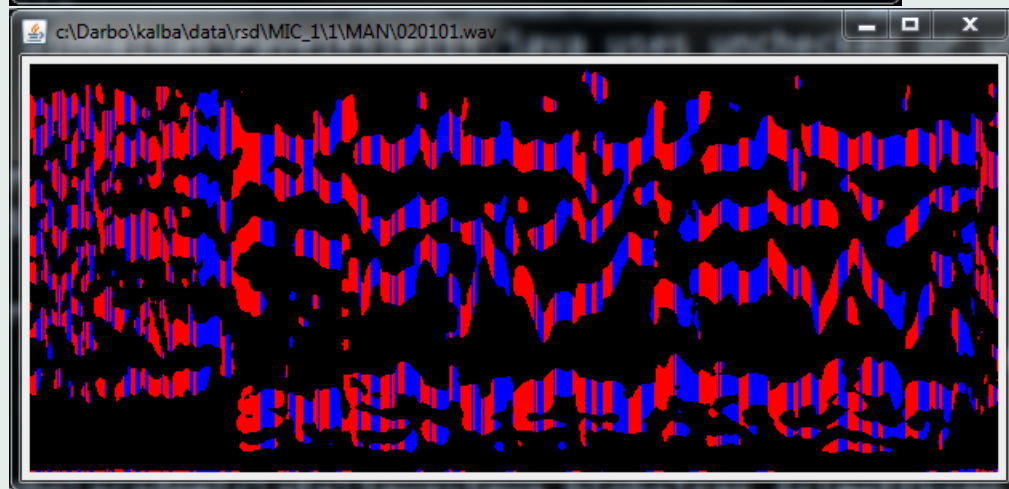
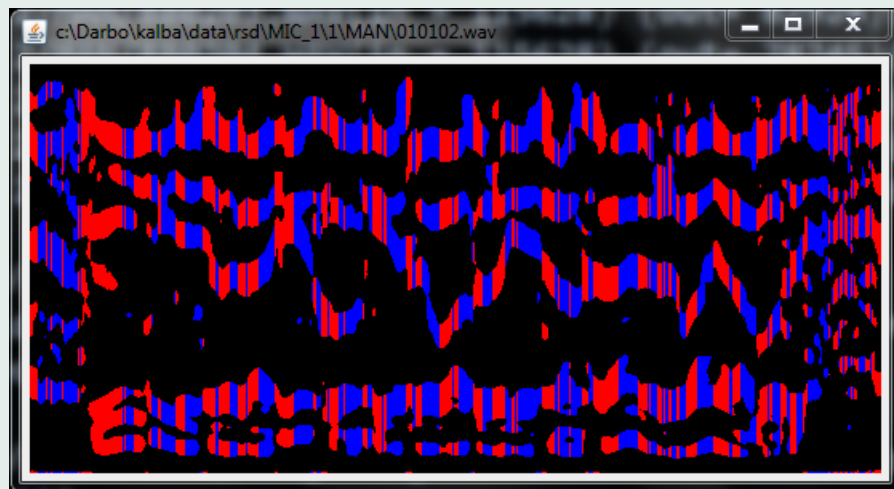
Page 8 of 19

Go Back

Full Screen

Close

Quit



5 pav.: 010102.wav ir 020101.wav balso duomenų binariniai požymiai

Home Page

Title Page

◀ ▶

◀ ▶

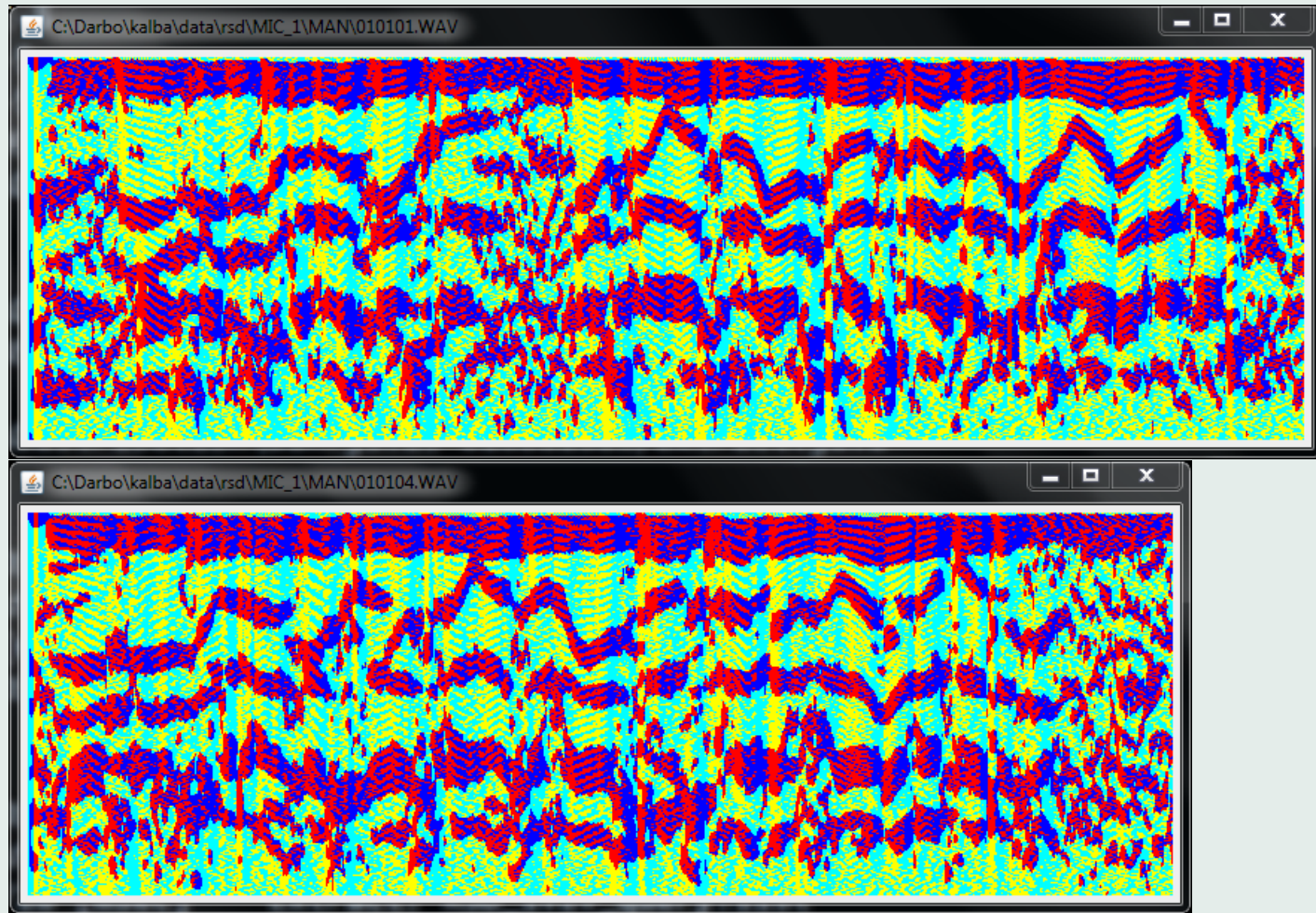
Page 9 of 19

Go Back

Full Screen

Close

Quit



6 pav.: 010101.wav ir 010104.wav balso duomenų kvadra požymiai

siūloma naudoti tokią techniką:

- 512 bitų vektorių supakuoti į $512/64 = 8$ long tipo masyvą;
- užpildyti 2^{16} ilgio 16-os bitų ilgio indeksų nulinių bitų kiekį. Pvz. $i = 0$ turi visus 16-a nulinių bitų, $i = 1$ - penkiolika, $i = 2$ - penkiolika, $i = 3$ - keturiolika, ..., $i = 2^{16} - 1$ - nulį bitų. Šių nulinių bitų skaitliuką užpildykite skaičiavimų pradžioją vieną kartą;
- dviejų long tipo supakuotų požymių $long\ a, b$ panašumą greitai įvertinkite tokiu būdu:

$$long\ ab = a\ xor\ b;$$

$$similarity(a, b) = zbc[ab \& 0xffff] + zbc[(ab >> 16) \& 0xffff] +$$

$$zbc[(ab >> 16) \& 0xffff] + zbc[(ab >> 32) \& 0xffff] + zbc[(ab >> 48) \& 0xffff].$$

Čia xor žymi binarinę XOR operaciją, $zbc[i]$ - i -ojo indekso pirmųjų 16-iolikos dvejetainės išraiškos nulinių bitų kiekis (zbc - "zero bits count").

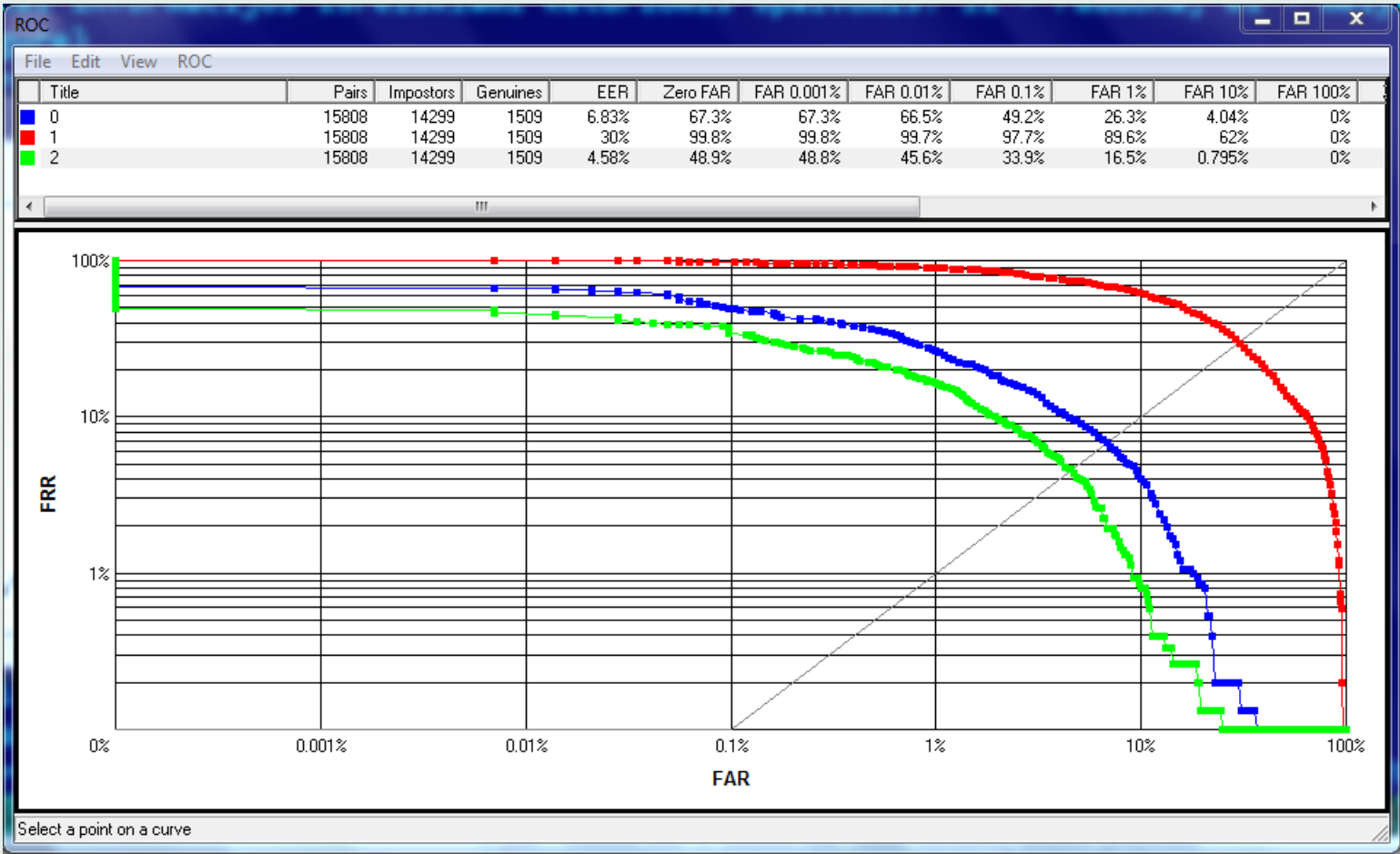
Tokiu būdu greitai apskaičiuosite spektrogramų dviejų stulpelių panašumą. Sulyginus visus dviejų $512 \times N$ ir $512 \times M$ dydžio spektrogramų stulpelius, gausite $N \times M$ panašumų matricą. Kad greičiau vykėtų skaičiavimai, galite naudoti pvz. ir vienos ir kitos spektrogramos tik kas trečią stulpelį ir taip sunaudosite apie dešimt kartų mažiau laiko.

Tolimesniam palyginimui naudokite DTW techniką analogišką parašų palyginimui. Paskaitoje bus pateiktos DTW modifikacijos, kurios pasiteisina lyginant balso pavyzdžius. 7 paveikslėlis iliustruoja dviejų rusiškų frazių palyginimo rezultatus. Laikome, kad dvi lyginamos frazės yra autentiškos (tinkamos), jei jos pasakytos vieno diktoriaus ir sutampa frazių tekstai. Mėlyna DET kreivė atitinka verifikacijos rezultatų kokybę naudojant tiesioginę DTW metriką, raudoną - jungtinę metriką, ir žalia - dviejų metrių svertinį vidurkį, imant tiesioginę metriką su svoriu $w_1 = 2$ ir jungtinę su neigiamu svoriu $w_2 = -1$. Įrašant rezultatus "*.roc" failo formatu, svertinis panašumas dauginamas iš 1000 ir apvalinama iki neneigiamo sveikąjo skaičiaus.

0.3. Praktinis darbas

Realizuokite fiksuoto teksto ir diktoriaus atpažinimo algoritmą. Bus vertinimas kiekvienas jūsų kuriamo balso atpažinimo algoritmo etapas. Algoritmo kūrimo etapai, jų vertinimo balai ir terminai pateikiami žemiau. Papildomai galutinis balso atpažinimo algoritmas bus vertinamas pagal pateiktų algoritmui nežinomų balso pavyzdžių atpažinimo kokybę. Kokybė vertinama pagal algoritmo DET kreivę, palyginant jūsų algoritmu gautą DET kreivę su kitų autorių (jūsų kolegų) gautomis DET kreivėmis.

Reikalingų algoritmui atiderinti balso pavyzdžius rasite čia. Balso atpažinimo užduotis yra pri-



7 pav.: Dviejų rusiškų frazių palyginimo rezultatai

valoma. Darbas atliekamas auditorijoje pratybų metu arba savarankiškai namuose atsiskaitant auditorijoje pratyboms skirtu laiku už atskirus etapus. Išimtiniais atvejais (pvz. esate ilgalaikėje studijų komandiruotėje), galima atsiskaityti *internetu* pagal nurodytus terminus siunčiant kodus ir vėliau juos atsiskaitant gyvai. Programavimo kalba neribojama, tačiau programuojantys java kalba turės privalumą, nes bus galimybė tiesiogiai naudotis pateikiamais java kodais. Taip pat galėsite naudotis užduočių, kurioms terminas baigėsi, java kodais atlikti dar nepradelstoms pagal terminus užduotims.

1. Balso duomenų nuskaitymas, spektrogramos ir formančių apskaičiavimas.

Nuskaitykite ir vizualizuokite nuskaitytus balso duomenis. Balso duomenims nuskaityti galite naudoti [WaveReader](#) java kodą.

Apskaičiuokite nuskaitytų balso duomenų spektrogramą ir įvertinkite spektrogramos formantes. Atliekant šią užduotį galite naudoti jums tinkamą [biblioteką](#) arba pasinaudoti [AsimetEkspo](#) java kodu.

Atsiskaitoma pademonstruojant nuskaitytų duomenų grafiką, spektrogramos ir formančių vaizdus.

Vertinama iki [0.75 balo](#), iki [10.25](#) ([visuomet imtinai](#), [bakalaurams diena vėliau](#)).

2. Balso požymių apskaičiavimas, jų binarizavimas ir vizualizacija.

Apskaičiuokite balso duomenų "*.wav" failo požymius. "*.wav" failų sąrašą imkite iš *.list failo. Balso požymių apskaičiavimui ir jų binarizacijai rekomenduojama naudoti **extract.bat** komandą. Nuskaitykite "*.sab" formato balso šablonus. "*.wav" failų sąrašą imkite iš *.list failo. Rekomenduojama nuskaitytą šabloną saugoti atskiroje *Model* klasėje, o visus *.list sąrašo failus *Model[]* masyve.

Vizualizuokite keletą nuskaitytų *Model* objektų binarinius požymius.

Atsiskaitoma pademonstruojant keleto nuskaitytų *Model* objektų binarinius požymių vaizdus. Binarinių požymių vizualizacijos kodo pavyzdį rasite **Extract.java** klasėje. *010101.sab* ir *010104.sab* binarinių požymių vizualizacijos pavyzdžiai pateikti 8 ir 9 paveikslėliuose. Jūsų gaunami paveikslėliai gali kiek skirtis, nes šių konspektų iliustracijos galėjo būti padarytos su kiek pakeistais *Extract.java* faile esančiais parametrais.

3. Balso požymių palyginimas.

Palyginkite balso duomenų sablonų poras. Failų sąrašą imkite iš *.list failų. Balso požymių apskaičiavimui ir jų binarizacijai rekomenduojama naudoti **extract.bat** komandą.

Atliekant užduotį rekomenduojama užpildyti **Match.java** klasės fiktyvias kodo eilutes (75-78 ir 99-108 *Match.java*) reikiamomis pagal fiktyvaus kodo aprašymą. DET kreivės apskaičiavimui

Home Page

Title Page

◀ ▶

◀ ▶

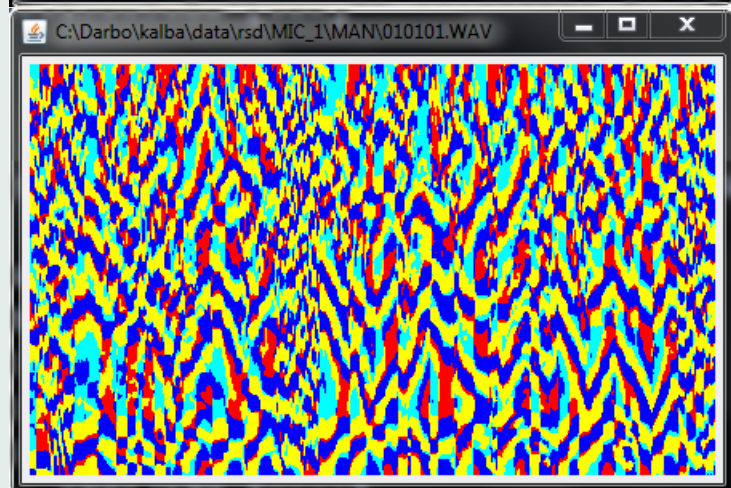
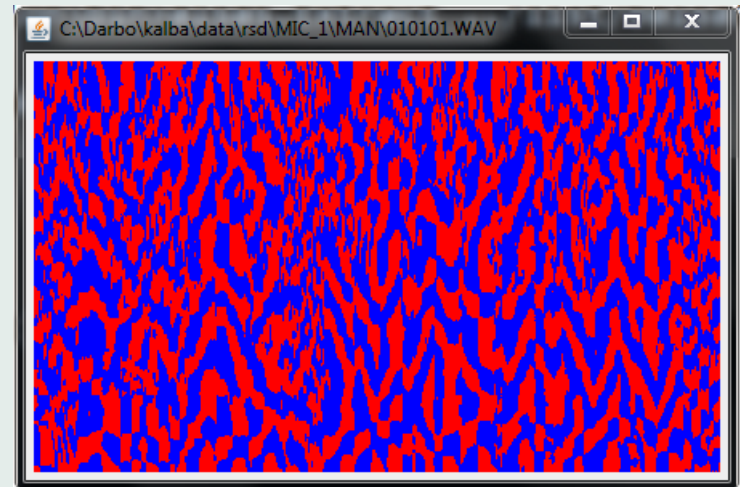
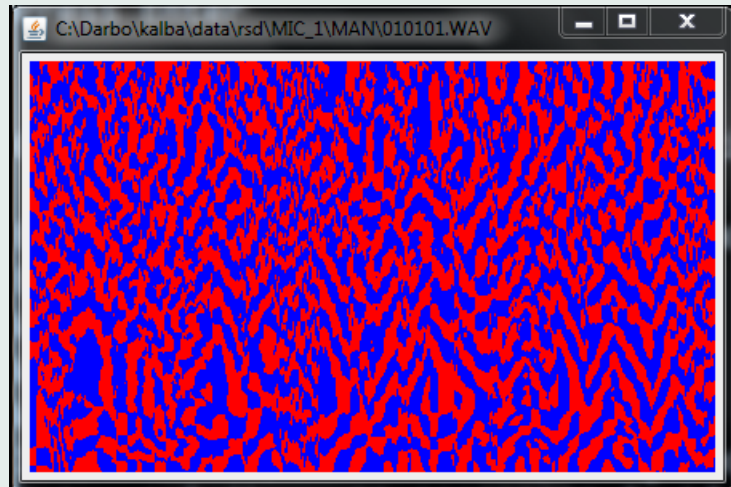
Page 15 of 19

Go Back

Full Screen

Close

Quit



8 pav.: Šablono 010101.sab binariniai požymiai

Home Page

Title Page

◀ ▶

◀ ▶

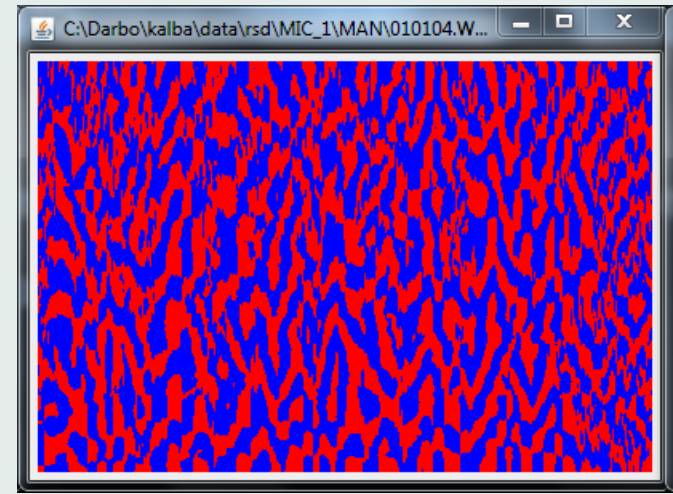
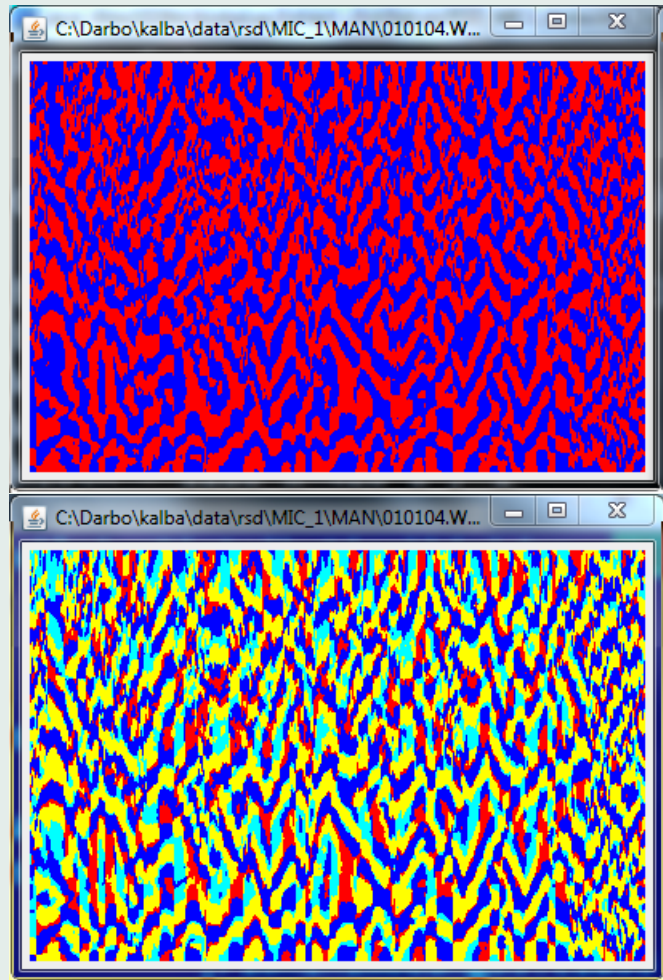
Page 16 of 19

Go Back

Full Screen

Close

Quit



9 pav.: Šablono 010104.sab binariniai požymiai

Home Page

Title Page

◀◀ ▶▶

◀ ▶

Page 17 of 19

Go Back

Full Screen

Close

Quit

rekomenduojama naudoti `match.bat` komandą.

Atsiskaitoma pademonstruojant analogus 10 paveikslėlio ir 11 DET kreivės, gautos lyginant *list/RSD_MIC1_mokymo_1v.list* ir *list/RSD_MIC1_testavimo_1v2v.list* sąrašų šablonų poras.

Vertinama iki 1 balo, atsiskaitoma iki 12.12.

Home Page

Title Page



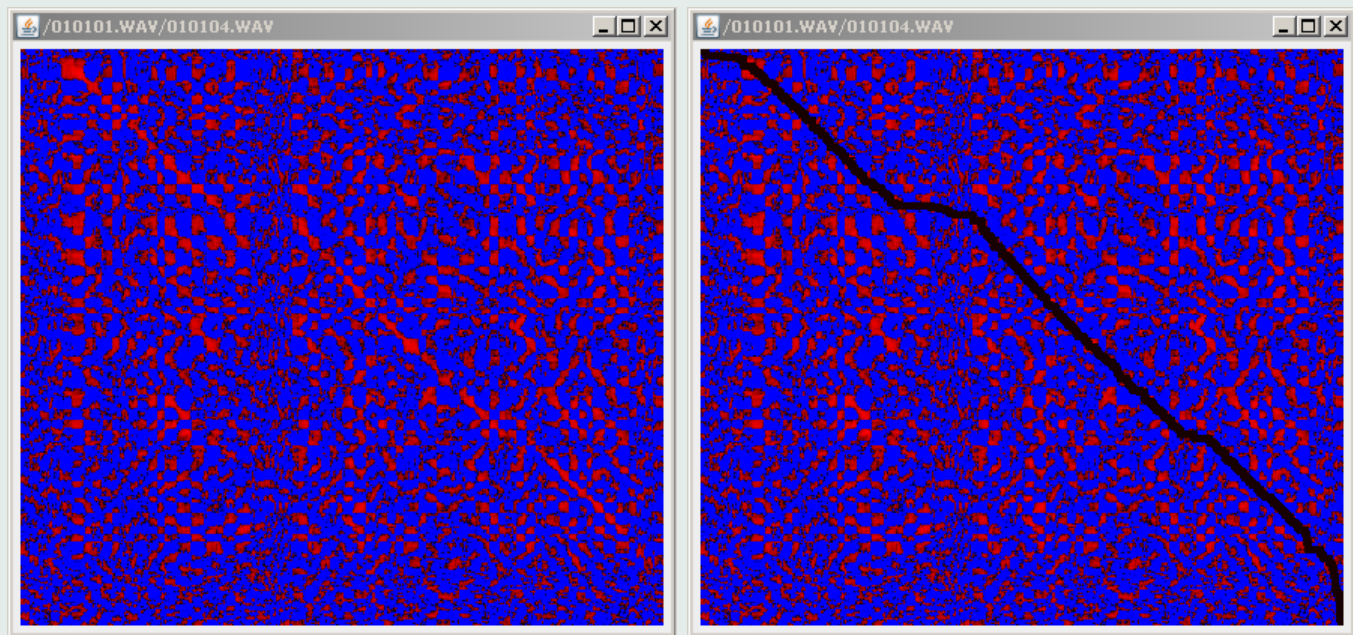
Page 18 of 19

Go Back

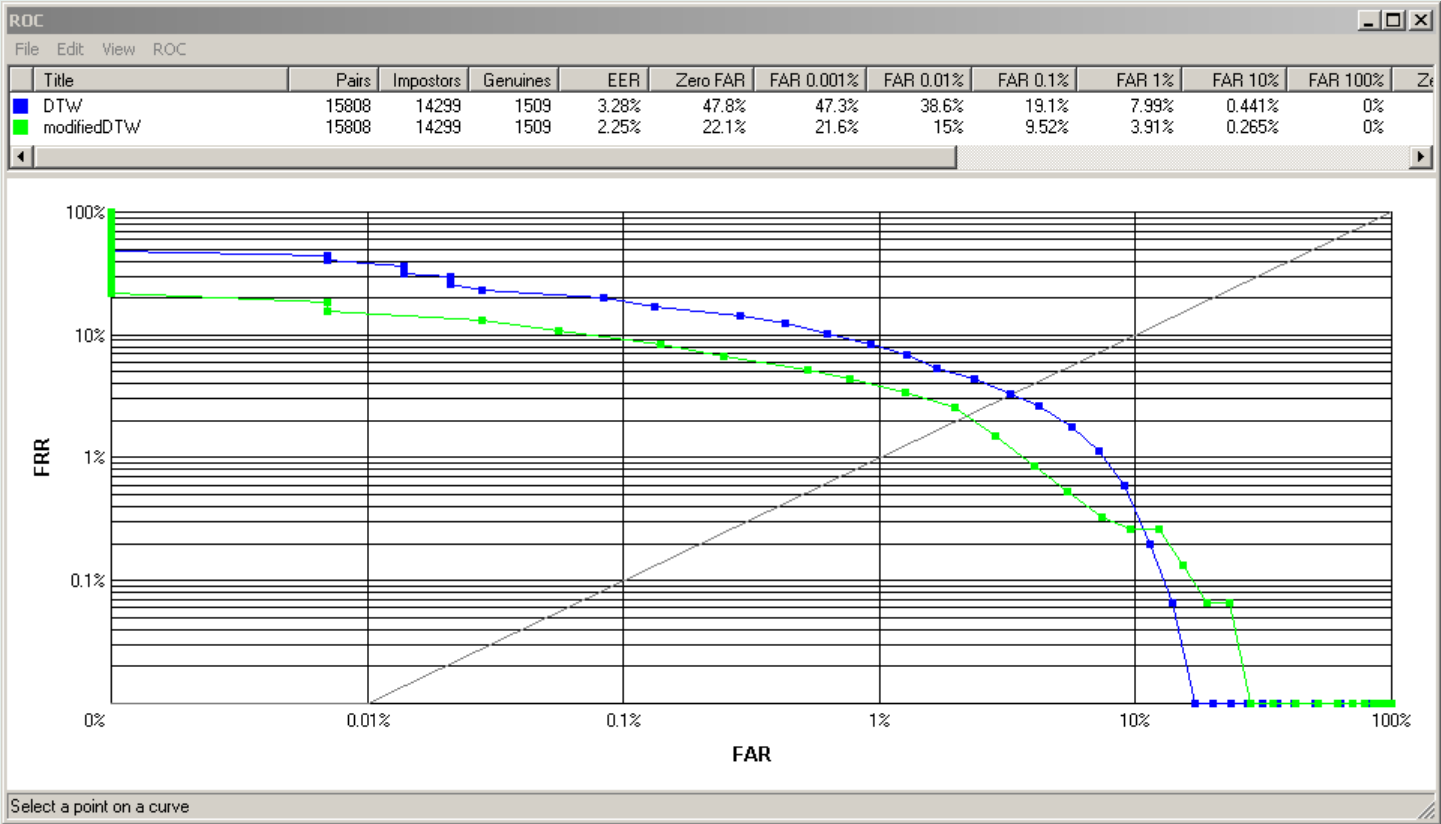
Full Screen

Close

Quit



10 pav.: 010101.sab ir 010104.sab kadrų panašumų matrica (kairėje) ir ta pati matrica su ištrintomis aplink DTW trajektoriją panašumų reikšmėmis (dešinėje).



11 pav.: Balso ir diktoriaus atpažinimo DET kreivės gautos lyginant RSD_MIC1_mokymo_1v.list ir RSD_MIC1_testavimo_1v2v.list sąrašų poras. Mėlyna kreivė žymi standartiniu DTW gautą DET kreivę, o žalia - kombinuojant standartinio DTW ir ištrintos pagrindinės DTW trajektorijos aplinkos panašumų reikšmes.