

**სამეტყველო ბგერების ოპტიკური გამოსახულებებით  
ამოცნობის ავტომატური სისტემა**

ოთარ ნატროშვილი, თემურ ბერიანიძე  
საქართველოს ტექნიკური უნივერსიტეტი

**რეზიუმე**

წარმოდგენილ სტატიაში შემუშავებულია სამეტყველო ბგერითი სახეების ამოცნობის იერარქიული სტრუქტურა მათი შესაბამისი ოპტიკური გამოსახულებების ადეკვატურ ოპტიკურ ეტალონურ გამოსახულებებთან ავტომატური შედარების გზით. მოცემულია ამოცნობის პროცედურების წარმოდგენა გრაფების დახმარებით.

**საკვანძო სიტყვები:** სამეტყველო ბგერები. ამოცნობის სისტემა. ოპტიკური გამოსახულებები.

**1. შესავალი**

სამეტყველო ბგერების, ისე როგორც ნებისმიერი სხვა აკუსტიკური სიგნალების ოპტიკური გამოსახულებებით ამოცნობის მეთოდოლოგიის შემუშავებისას უპირველეს ყოვლისა ისმება კითხვა: საჭიროა თუ არა სამეტყველო ბგერების ავტომატურად გამომცნობი მოწყობილობის (სისტემის, კომპლექსის) აგებისას მივყვეთ აკუსტიკური სიგნალების სმენითი აღქმის მოდელს?, თუ ისინი უნდა დაშუშავდნენ დამოუკიდებლად, აუვლით რა გვერდს ბიონიკას? საკითხი ბიონიკის როლის შესახებ, ე.ი. იმ ცოდნის შესახებ, რომელიც ძირითადად სწავლობს აღქმის ბიოლოგიურ მექანიზმს და ბიოლოგიური მონაცემებიდან ახალი იდეების გადმოღებას, თანამედროვე მეცნიერებაში კვლავ ბურუსითაა მოცული (მითუმეტეს ამ იდეების პრაქტიკულ გამოყენებაზე ჯერ-ჯერობით ლაპარაკიც კი ნაადრევია). საკითხის სირთულიდან გამომდინარე ამ კუთხით ჯერ კიდევ ბევრი რამ გაუგებარი და შეუსწავლელია. იმისათვის, რომ ბიონიკამ შეძლოს დაეხმაროს ინჟინერ-პრაქტიკოსებს სამეტყველო ბგერების ობიექტური ამოცნობის მეთოდების შემუშავებისას, საჭიროა სმენითი აღქმის ბიოლოგიური მექანიზმის შეძგომი შესწავლა. პრაქტიკოს-ინჟინრები ფიზოლოგებთან ერთად ფონემების ამოცნობის მიზნით აწყდებიან მრავალ სიძნელეებს. ასეთი სიძნელეების გადალახვა მნიშვნელოვან წილად შესაძლებელი გახდება თუ კი თანამედროვე ტექნიკით (უპირველეს ყოვლისა კომპიუტერების გამოყენებით) რეალიზებული იქნება სამეტყველო ბგერების ავტომატური ამოცნობის მეთოდები და საშუალებები უფრო ფართო მასშტაბებით, ვიდრე ეს ამჟამად შეინიშნება (და არ შემოიფარგლება მხოლოდ რამოდენიმე სიტყვის კრებულით).

ამ სიტუაციიდან გამოსვლა ალბათ შეიძლება თუ ბგერებს გადავაქცევთ ხილულ ობიექტებად (ჩვენს შემთხვევაში ოპტიკურ გამოსახულებებად) და გამოვიყენებთ გამოსახულების ამოცნობის მეთოდებს, რომლებიც დღეს-დღეობით უფრო დახვეწილია, ვიდრე საკუთრივ ბგერების (ფონემების) ამოცნობის მეთოდები. ყოველივე ამაში დიდ იმედებს იძლევა თანამედროვე კომპიუტერული ტექნიკა და მისი გაფართოებული ფუნქციონალური შესაძლებლობები. კომპიუტერული ტექნიკის განვითარებამ ხილული გამოსახულებების ურთიერთ შედარებით შესაბამისი ფონემების ამოცნობის საქმეში შეიძლება დაგვაყენოს ამ მიზნის მიღწევის უმოკლეს გზაზე. პრაქტიკა ამას ნაწილობრივ ადასტურებს კიდევ. აქვე კვლავ შევნიშნოთ, რომ სამეტყველო ბგერების ავტომატური ამოცნობის სფეროში მრავალი წარმატება მიიღწევა არა ბიონიკის რეკომენდაციების წყალობით, არამედ ინჟინრების მიერ ტექნიკის სფეროში ისეთი ცნობილი საშუალებების გამოყენებით, რომლებიც ნაკლებადაა (ანდა სრულებით არაა) შეღწეული ფიზოლოგიისა და ფსიქოლოგიის საკითხებში. პირიქით, შეიმჩნევა, ტენდენცია, რომ სამეტყველო ბგერების ამოცნობის ტექნიკა ამჟამად უფრო ვითარდება სუფთა ტექნიკური კუთხით, ე.ი. საინჟინერო მიმართულებით (გამოცნობის პროცესში ფიზიოლოგიისა ფსიქოლოგიური მომენტების გამოყენების გარეშე). აღნიშნულ მიდგომას ანვითარებს წარმოდგენილი სტატიის კვლევის თემატიკაც – ბგერითი სახეების ოპტიკური გამოსახულებებით ამოცნობის მეთოდოლოგიის სრულყოფა და მათი რეალიზაციის ტექნიკური საშუალებების შემუშავება.

წარმოდგენილ სტატიაში ყურადღება გამახვილებულია სწორედ ზემოთ ნახსენები კუთხით. კერძოდ, გამოკვლეულია ოპტიკური გამოსახულებებით ბგერების ტექნიკური საშუალებებით ამოცნობის ზოგადი მეთოდოლოგიური სქემა.

ამ მიზნით აღნიშნულ სტატიაში შემუშავებული და განხილულია ოპტიკური გამოსახულებებით სამეტყველო ბგერების ავტომატური ამომცნობი მოწყობილობის აგების ბლოკ-სქემა, დამუშავებულია ბგერითი ფონემების ავტომატური ამომცნობის იერარქიული მიდგომა, ნაჩვენებია ოპტიკური გამოსახულებებით ბგერითი სახეების ამომცნობის პროცედურების წარმართვის მაგალითი გრაფების დახმარებით.

## **2. ძირითადი ნაწილი**

ბგერითი სახეების (ფონემების) ავტომატური გარდაქმნა და მათი ამომცნობა ტექნიკური საშუალებებით, მეტად აქტუალურია თანამედროვე მეცნიერებისა და ტექნიკის, მათ შორის მართვის ავტომატიზებული სისტემების მრავალი ამოცანის გადასაწყვეტად. ეს ამოცანები პირველ რიგში დაკავშირებულია ტექნიკური მოწყობილობებით (უირველეს ყოვლისა კომპიუტერული ტექნიკის დახმარებით) სინთეზური მეტყველების მეთოდებისა და საშუალებების გამოყენებასთან, ტექნიკური ხმაურით მოწყობილობების, დაზვა-დანადგარების, ენერგეტიკული აგრეგატების და ა.შ. მუშაუნარიანობის დიაგნოსტიკისათვის. ტექნიკურ ამომცნობათა ეფექტური გადაწყვეტა მეტად პრობლემატურია, აგრეთვე ადამიანის სალაპარაკო ენით კომპიუტერთან ურთიერთობისათვისაც. ამისათვის კი საჭიროა ბგერების (როგორც სამეტყველო, ასევე ტექნიკური ბგერების) გარდაქმნის ახალი, უფრო ეფექტური მეთოდებისა და საშუალებების დამუშავება, რაც თავის მხრივ მოითხოვს აღნიშნულ საკითხებთან დაკავშირებული პრობლემების ღრმა და საფუძვლიანი გამოკვლევების ჩატარებას.

წარმოდგენილ სტატიაში შემოთავაზებული ახალი მიდგომის – ფონემების (აკუსტიკური, მათ შორის სამეტყველო ბგერების) შესაბამისი ოპტიკური გამოსახულებებით ამომცნობის კუთხით შეიძლება გამოვკვეთოდ ამომცნობის ურთიერთდაკავშირებული შემდეგი ორი მეთოდოლოგიური სქემა:

1. მეთოდოლოგიური სქემა ამოსაცნობი ფონემის ფიზიკური წარმოდგენის თვალსაზრისით;
2. მეთოდოლოგიური სქემა ამომცნობის პროცესის წარმართვის ტექნიკური რეალიზაციის თვალსაზრისით.

ფიზიკური წარმოდგენის თვალსაზრისით ფონემების ოპტიკური გამოსახულებების ნებისმიერი სიმბოლიკა ადეკვატურად უნდა შეესაბამებოდეს აკუსტიკური სახის (ფონემის) აღწერას და ამ კუთხით დიდი როლი მიეკუთვნება სათანადო სიზუსტის ამსახველი (წარმოდგენი) ტექნიკური მოწყობილობების შექმნას.

ამომცნობის პროცესის ტექნიკური რეალიზაციის თვალსაზრისით ფონემების ოპტიკური გამოსახულებებით ამომცნობისათვის საჭიროა: ბგერა-მიმღები (მაგალითად, მიკროფონი გამაძლიერებელთან ერთად); ფონემების ექსტრაქტორი (ხელსაწყო, რომლის დახმარებით შესაძლებელი გახდება ფონემების დინამიური გამოყოფა (ამოფესვა) მეტყველების ბგერითი თანამიმდევრობიდან); დინამიური ანალიზატორი მონაცემების (ამჯერად უკვე ფონემების შესაბამისი ოპტიკური გამოსახულებების) კომპიუტერში შესატანად და ბოლოს, გამომთვლელი მანქანა (პერსონალური კომპიუტერი ან საკმარისი სიმძლავრის მენიფრემი ეტალონთან მსგავსების შესახებ საბოლოო გადაწყვეტილების ამოსატანად) [1-3].

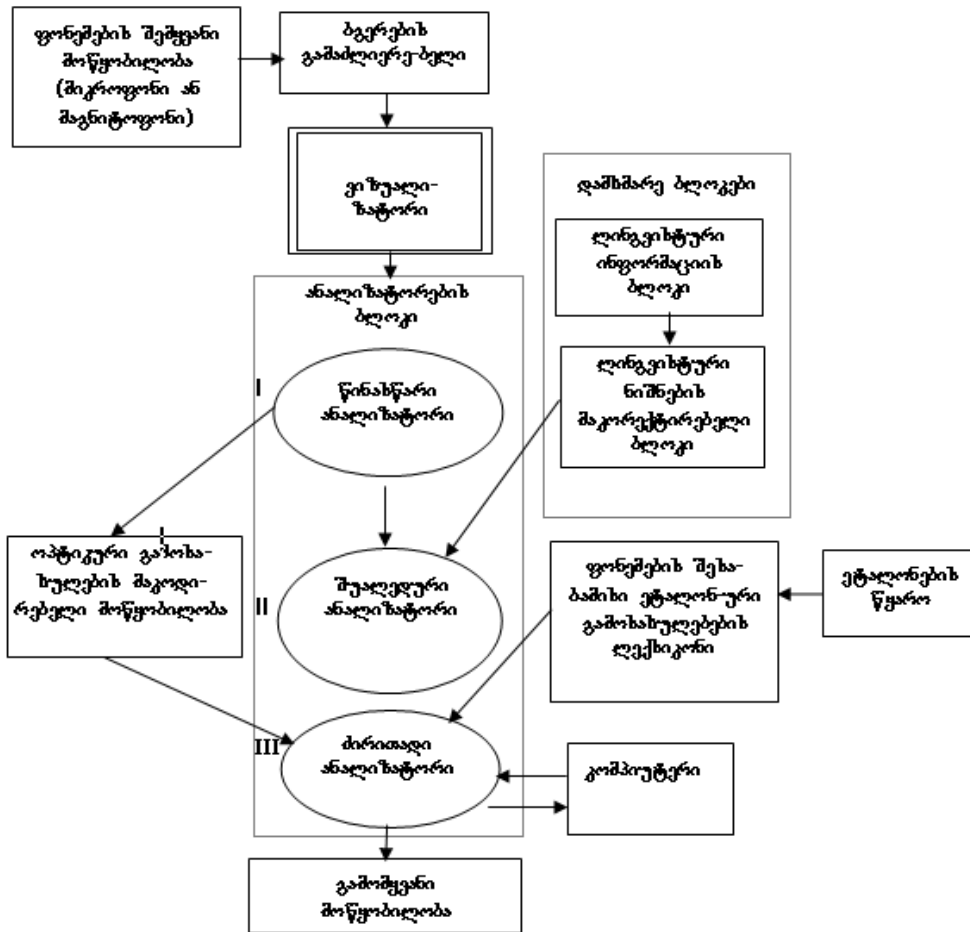
ოპტიკურ გამოსახულებებად ბგერების გარდაქმნის შემდეგ ამომცნობი მოწყობილობის სხვადასხვა რგოლში ფართოდ და ეფექტურად უნდა იქნეს გამოყენებული ოპტიკურ სახეთა ეტალონურ გამოსახულებებთან შედარებისა და ამომცნობის თანამედროვე მეთოდები და ამ მეთოდების სარეალიზაციოდ განკუთვნილი გამოთვლითი ტექნიკის ამჟამად არსებული აპარატურულ-პროგრამული საშუალებები.

ოპტიკური გამოსახულებებით ბგერითი სახეების ამომცნობი მოწყობილობების (სისტემების, კომპლექსების) ზოგადი ბლოკ-სქემა, რომელიც დაფუძნებულია ზემოთ ფორმულირებულ მეთოდოლოგიურ მიდგომებზე, ნაჩვენებია ნახ.1-ზე.

1-ელ ნახაზზე გამოსახულ მოწყობილობებში ამომცნობის მეთოდის სპეციფიკიდან გამომდინარე, საჭიროებისდამიხედვით შეიძლება შეტანილი იქნეს დამატებითი ბლოკებიც (მათ შორის სპეცდანიშნულების ფუნქციონალური კვანძებიც), რომლებიც ამ სქემას მისცემს უფრო რთულ, იერარქიულ სახეს. დამატებით კვანძებს მიეკუთვნება პირველ რიგში ლინგვისტიკური ინფორმაციისა და ლინგვისტიკური ნიშნების მაკორექტირებელი ბლოკები.

ანალიზატორების ბლოკში შემავალი წინასწარი ანალიზატორის (I) ძირითადი დანიშნულებაა გარკვიოს გამოსაცნობი სახეები წარმოდგენენ თუ არა სამეტყველო ბგერებს ან მანქანის ხმაურს; შუალედური ანალიზატორი (II) აანალიზებს ძირითადად ანალიზატორისაკენ გასავაზვნ ინფორმაციას

ლინგვისტური ნიშნების გამოყენების საჭიროებას, ზოლო მესამე ძირითად ანალიზატორში (III) სწარმოებს ძირითადი ოპერაციები მიმდინარე გამოსაცნობი ბგერების შესაბამისი (ადექვატური) ოპტიკური გამოსახულებების ეტალონურ გამოსახულებებთან შესადარებლად და გამოცნობის (ან არგამოცნობის) შესახებ გადაწყვეტილებების მისაღებად. ხშირად III – ანალიზატორი და კომპიუტერი შეიძლება განხილული იქნენ ერთ მთლიან ფუნქციონალურ ბლოკად.

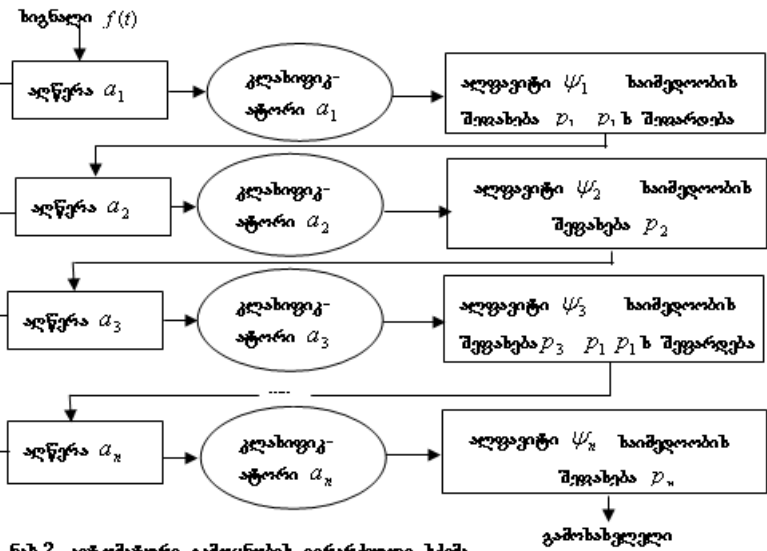


ნახ.1. ოპტიკური გამოსახულებებით ბგერების გამომცნობი მოწყობილობის ბლოკ-სქემა

განვიხილოთ ფონემების გამოცნობის იერარქიული სქემა, რომელიც მოცემულია მე-2 ნახაზზე. ამ სქემის მიხედვით შეიძლება ავლნიშნოთ ოპტიკური გამოსახულებებით მათი (სამეტყველო ბგერითი სახეების, ანუ ფონემების) გამოცნობის შემდეგი იერარქიული პროცედურა: შესასვლელზე მიეწოდება ამოსაცნობი  $f(t)$  სიგნალი რაღაც დროითი თანამიმდევრობით. პირველ დონეზე წარმოებს გადასვლა უფრო მარტივ  $a_1$ - ალწერაზე. რეალიზაციებს, რომლებიც მიღებული იქნება ამ დროს, შეიძლება ჰქონდეს სონოგრამების ხასიათი, საიდანაც შეიძლება ამოირჩეს (ამოიფესვოს) რაიმე ელემენტები – სემენტები. ამგვარად  $f(t)$  სიგნალის გადაშუქების შედეგად მიიღება სემენტების კრებული. გამოცნობა პირველ იერარქიულ საფეხურზე შემოიფარგლება მხოლოდ სემენტების კლასიფიკაციით. ამასთან თითოეულ საფეხურზე ნაპოვნ უნდა იქნას ასევე ამ ოპერაციის საიმედოობის სიდიდეები.

მეორე იერარქიულმა საფეხურმა უნდა გამოიტანოს დასკვნა სიგნალის წყაროს ხასიათთან მიმართებით, ან თუ ეს შესაძლებელია, ამავე საფეხურზე განისაზღვროს ამოსაცნობი სახე. ამ საფეხურზე გადაწყვეტი ფუნქციების როლში შეიძლება იყოს გამოსახულების სხვადასხვა სემენტების ალბათობების განაწილებათა პოვნის ალგორითმი. ამ საფეხურზე, გადაწყვეტილების შედარების გზით ეტალონურ განაწილებასთან, კლასიფიკატორი საშუალებას იძლევა გაკეთდეს დასკვნა სიგნალის წყაროს ამა თუ იმ სახეზე მიკუთვნების შესახებ. ასევე წარმოებს ამავე საფეხურზე გაკეთებული დასკვნების საიმედოობის შეფასება. მომდევნო დროით ეტაპებზე რეალიზდება დანარჩენი საფეხურები, სანამ არ მიიღწევა ეტალონთან საბოლოო მსგავსება მაქსიმალური საიმედოობით, რის შემდეგაც გამოსახულებზე გაიცემა დასტურის სიგნალი ამოსაცნობი ფონემის ეტალონთან მსგავსობის

(ან არამსგავსობის) შესახებ. ამგვარად, ვინაიდან წყაროს თითოეული ტიპისათვის არსებობს რეჟიმების სხვადასხვა სახე, რომლებიც შეიძლება მიეკუთვნოს ასევე სხვადასხვა კლასებს, შეიძლება საჭირო გახდეს შემდეგი – ამოცნობის მომდევნო საფეხური, სადაც უფრო ზუსტდება სახეთა კლასიფიკაცია მათი კონკრეტული ნიშნების მიხედვით. თუმცა დაბრკოლებას წარმოადგენს ის, რომ შესაძლოა ვარიანტების რიცხვი წარმოიქმნას ძალზე მაღალი, ამიტომ შეიძლება დაგვიჭირდეს იერარქიის სხვა, უფრო მაღალი საფეხურებიც, სადაც თანდათან დაკონკრეტდება ამოსაცნობ სახეთა დამახასიათებელი ნიშნები. ამგვარად, შეიძლება გავყვეთ უფრო და უფრო ალბათურ ვარიანტებს სხვადასხვა შეზღუდვების გამოყენებით და ბოლო საფეხურზე დადგინდება საბოლოო დეტალური სქემა, რომლის მიხედვითაც გამოითანება ობიექტური გადაწყვეტილება სახეთა გამოცნობის შესახებ რისკის მინიმალური მნიშვნელობით.



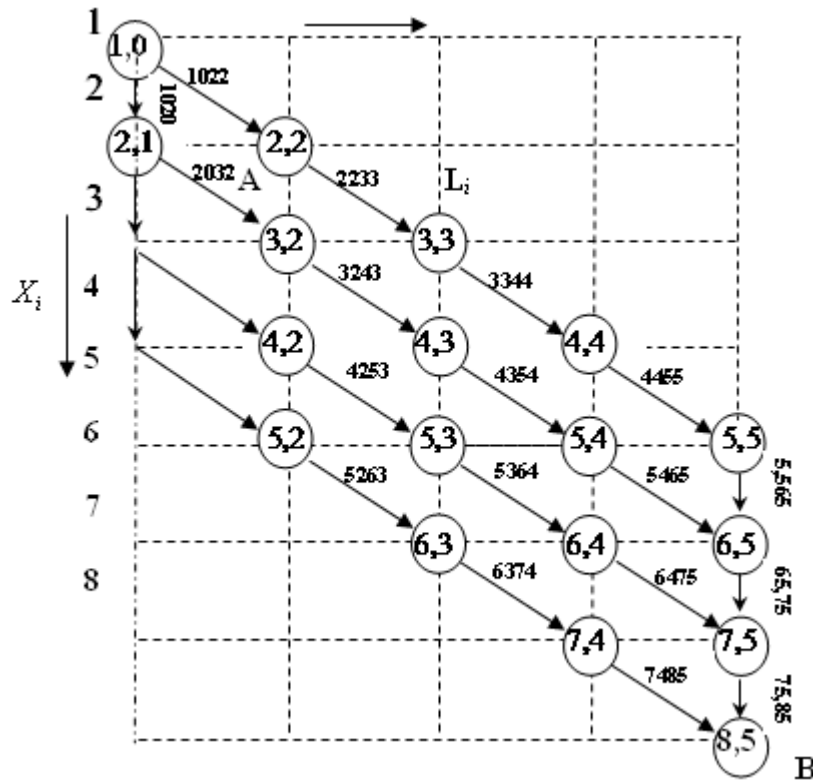
ნახ.2. ავტომატური გამოცნობის იერარქიული სქემა

რთულ მოდელებში ჩაირთვება სპეციალური მოწყობილობები (ე.წ. „სუპერვაიზერები”, რომელთა დახმარებითაც შესაძლებელი გახდება წარმართოს გადაწყვეტილების მიღების პროცესის მართვა თითოეულ იერარქიულ საფეხურზე.

თეორიული კუთხით შესაძლებელია ვაჩვენოთ აკუსტიკური სიგნალების ეტალონური გამოსახულებებით გამოცნობის პროცედურების წარმართვის გეომეტრიული ინტერპრეტაცია. ამოცნობის ამოცანა მდგომარეობს ყველა ეტალონიდან ამოსაცნობ გამოსახულებასთან (X-თან) შედარების გზით უფრო ახლო მდგომი ეტალონის (ე.ი. შესაბამისი (ადექვატური) ოპტიკური ეტალონური გამოსახულების) მოძებნაში და მის პოვნაში. ამისათვის საჭიროა შევარჩიოთ ხერხი, რომლის მიხედვითაც უნდა ავაგოთ ეტალონური სიმრავლე და გარკვეული წესით წარვმართოთ ძებნის განშტოებადი პროცესი. ასეთი პროცესის გრაფებით ასახვის ერთ-ერთი მაგალითი ნაჩვენებია მე-3 ნახაზზე. იგი დაწვრილებით განხილულია [2]-ში.

გამოცნობის კრიტერიუმები შეიძლება შემოწმებული იქნენ ანალიტიკურად შემდეგი გამოსახულებით:

$$\min_k \left\{ \min_{E_\ell^k \in \Omega_\ell^k} \left| X_\ell - E_\ell(E_q^k) \right|^2 \right\} = \min_k \left\{ \min_{E_\ell^k \in \Omega_\ell^k} \sum_{i=1}^{\ell} \left| X_\ell - \ell_{j(i)} \right|^2 \right\} \quad (1)$$



ნახ.3. ოპტიკური გამოსახულებებით ბგერითი სახეების გამოცნობის პროცედურების წარმოდგენა გრაფების დახმარებით

გამოსახულება (1) გამოსატავს ამოსაცნობი სახის  $X_e$  თანმიმდევრობის განსხვავებას ეტალონ-გამოსახულებასთან  $E_e$  და ამოცნობის ეფექტურობის ამოცანა დაიყვანება ამ სხვაობის მინიმუმის პოვნაზე, ე.ი. მაქსიმალურ მსგავსებამდე უმოკლესი გზის პოვნაზე.

### 3. დასკვნა

წარმოდგენილი სტატიის ერთ-ერთი მთავარი მიზანი ოპტიკური გამოსახულებებით ბგერითი სახეების (ფონემების) ავტომატური ამოცნობის საერთო მეთოდოლოგიური მიდგომების დამუშავებაა, სამეტყველო ან სხვა სახის აკუსტიკური (ბგერითი) სიგნალების გარდაქმნისა და ამოცნობის შესაბამისი ტექნიკური მოწყობილობების ეფექტური სტრუქტურების შესაქმნელად.

სამეტყველო აკუსტიკური ბგერების ამოცნობისას, შესაბამისი ოპტიკური გამოსახულებებით, ძალზე მნიშვნელოვანია ჯერ ერთი ფონემების ადეკვატური გარდაქმნა-წარმოდგენა ოპტიკური გამოსახულებებით და მეორე – არანაკლებ მნიშვნელოვანია ავტომატური ამოცნობისათვის სისტემაში ეტალონური ოპტიკური გამოსახულებების თანმიმდევრობების ეფექტური აგების, ე.ი. ამოცნობის პროცესის წარმართვის ეფექტური წესის შემუშავება, რომელიც შეიძლება მოითხოვდეს გამოსაცნობი სახისა და ეტალონების ურთიერთ შედარების პროცედურების დიდი რაოდენობის შემცველ სივრცეს. ისინი (ეტალონური გამოსახულებები) შეიძლება წინასწარ შეტანილი იქნენ კომპიუტერის მეხსიერებაში (ე.ი. ეტალონურ მონაცემთა ბაზაში) და კომპიუტერის პროცესორის დახმარებით გაანალიზდეს ამოცნობის ამა თუ იმ პროცესში როგორც მათი გამოყენების თანმიმდევრობა (ჯაჭვი), ასევე მათი გამოყენების რაოდენობა და საჭიროების ხარისხი.

### ლიტერატურა:

1. ბერიანიძე თ., ნატროშვილი ო., კამკამიძე კ. სამეტყველო ბგერების ვიზუალიზაციისა და ოპტიკური გამოსახულებებით მათი ამოცნობის მეთოდოლოგია. სტუ. შრ.კრ. №1(451). 2004

2. ბერიანიძე თ. ოპტიკური გამოსახულებებით ბგერითი სახეების გამოცნობის პროცედურათა წარმოდგენის მეთოდები გრაფების გამოყენებით. „მეცნიერება და ტექნოლოგიები“. საქ. მეცნ.აკად. ჟურნ. №10-12. თბ., 2003

3. Берианидзе Т.О., Натрошвили О.Г. Методы и средства распознавания звуковых образов оптическими изображениями фоном. Georgian Engineering News № 1. Тб., 2004.

### **THE AUTOMATIC SYSTEM OF RECOGNITION OF SPEECH SOUNDS WITH OPTICAL IMAGES**

Natroshvili Otar, Berianidze Temur  
Georgian Technical University

#### **Summary**

A hierarchical structure of recognition of speech sound images by automatic comparison of their corresponding optical images with adequate optical reference images is devolved. Representation of procedures of recognition by means of counts is given.

### **АВТОМАТИЧЕСКАЯ СИСТЕМА РАСПОЗНАВАНИЯ РЕЧЕВЫХ ЗВУКОВ ОПТИЧЕСКИМИ ИЗОБРАЖЕНИЯМИ**

Натрошвили О.Г., Берианидзе Т.О.  
Грузинский Технический Университет

#### **Резюме**

Разработана иерархическая структура распознавания речевых звуковых образов путем автоматического сравнения их соответствующих оптических изображений с адекватными оптическими эталонными изображениями. Дано представление процедур распознавания при помощи графов.