

ფურცელი და ჩეხიშევის გარდაქმნების ეფექტურობის შედარება და მათი პრაქტიკული
გამოყენება Text-To-Speech და Speech-To-Text ამოცანებში

სერგეი კაზარიანი

სამაგისტრო ნაშრომი წარდგენილია ილიას სახელმწიფო უნივერსიტეტის ბიზნესის,
ტექნოლოგიისა და განათლების, ფაკულტეტზე მათემატიკის მაგისტრის აკადემიური
ხარისხის მინიჭების მოთხოვნის შესაბამისად

თანამედროვე მათემატიკის ძირითადი პარადიგმები

სამეცნიერო ხელმძღვანელი: გიორგი რაქვიაშვილი

ილიას სახელმწიფო უნივერსიტეტი

თბილისი, 2024

განაცხადი

როგორც წარდგენილი სამაგისტრო ნაშრომის ავტორი, ვაცხადებ, რომ ნაშრომი წარმოადგენს ჩემს ორიგინალურ ნამუშევარს და არ შეიცავს სხვა ავტორების მიერ აქამდე გამოქვეყნებულ, გამოსაქვეყნებლად მიღებულ ან დასაცავას წარდგენილ მასალებს, რომლებიც ნაშრომში არ არის მოხსენიებული ან ციტირებული სათანადო წესების შესაბამისად

სერგეი კაზარიანი



თარიღი 15.06.2024

გამოყენებული ტერმინები

<p>გამოთვლის სირთულე $O(\dots)$</p>	<p>ალგორითმის გამოთვლის სირთულის შეფასება, რომელიც დამოკიდებულია ალგორითმში შემავალი მონაცემების რაოდენობაზე. მაგალითად $O(n)$ სირთულე აღნიშნავს, რომ ალგორითმის სირთულე იზრდება წრფივად შემავალი მონაცემების ზრდასთან ერთად; მონაცემების გაორმაგება იწვევს ორჯერ მეტ დროს ალგორითმის შესასრულებლად. $O(\log(N))$ - ლოგარითმული დრო, $O(N \log(N))$ - ლოგარითმეტიკული დრო, $O(N^2)$ - კვადრატული დრო.</p>
<p>Sample / სემპლი</p>	<p>უწყვეტი (ანალოგური) სიგნალის ანათვალი; დროის მომენტში რაოდენობრივი ინფორმაცია სიგნალის მახასიათებლების შესახებ.</p>
<p>Sampling Rate / დისკრეტიზაციის სიხშირე</p>	<p>ანათვლების (სამპლების) წაკითხვა წამში. აუდიო სიგნალის შემთხვევაში დისკრეტიზაციის სიხშირე უდრის 44100 ან 48000 Hz.</p>
<p>DFT – Discrete Fourier Transform</p>	<p>ფურიეს დისკრეტული გარდაქმნა იძლევა საშუალებას წარმოვადგინოთ დისკრეტული პერიოდული ფუნქცია სასრული სიხშირეების სახითა, სადაც თითოეულ სიხშირეს აქვს განსაზღვრული ამპლიტუდისა და ფაზის რიცხვითი მაჩვენებელი. განსაზღვრის თანახმად, ფურიეს დისკრეტული გარდაქმნა გვაძლევს ფუნქციის სპექტრალური წარმოდგენა.</p>
<p>FFT – Fast Fourier Transform</p>	<p>DFT - ს გამოთვლის ეფექტური ალგორითმი, რომელსაც ამცირებს გამოთვლის სირთულეს $O(N^2)$ - დან $O(N \log(N))$ - მდე</p>
<p>DCT – Discrete Chebyshev Transform</p>	<p>ჩებიშევის პოლინომებზე დაფუძნებული დისკრეტული გარდაქმნა.</p>

აბსტრაქტი

მეტყველების ამოცნობის მათემატიკური მეთოდები კვლავაც წარმოადგენენ მნიშვნელოვან და დიდი პოტენციალის მქონე დარგს. გამოთვლითი სიჩქარის ზრდამ შესაძლებელი გახადა აღნიშნული მეთოდების ფართოდ გამოყენება. სპექტრალური ანალიზი წარმოადგენს ერთ-ერთ საფეხურს მეტყველების ამოცნობის პროცესში, კერძოდ, იგულისხმება ხმოვანი სიგნალის სიხშირული სპექტრის სახით წარმოდგენა და მისი ანალიზი, რომელიც დაფუძნებულია ფურიეს სწრაფ გარდაქმნაზე (FFT). მიუხედავად FFT - პოპულარობისა, პრაქტიკული გამოყენების თვალსაზრისით, მისი რეალიზება დაკავშირებულია დიდ გამოთვლით რესურსებთან, განსაკუთრებით მაშინ, როცა ოპერაციაში მონაწილეობენ კომპლექსური რიცხვები. სამაგისტრო ნაშრომში შედარებულია FFT - სა და ჩებიშევის DCT გარდაქმნების ალგორითმების ეფექტურობა და შესრულების დრო. გარდა ამისა, შემოთავაზებულია ხმოვანი და თანხმოვანი ბგერების ეფექტური ანალიზი 256 დისკრეტული ორთონორმალური ჩებიშევის პოლინომების გამოყენებით. მარტივ მაგალითზე დაფუძნებით დადგინდა, რომ 256 DCT - ს საშუალებით შესაძლებელია ანალოგური ხმოვანი სიგნალის დროის არედან სიხშირულ არეში გადაყვანა, რაც მეტყველების ამოცნობის პროცესში აუცილებელ წინაპირობას წარმოადგენს. ამასთან ერთად 256 DCT - ს გამოყენებით შეგვიძლია მივიღოთ ერთდროულად ოთხი ფორმანტა F_1, F_2, F_3, F_4 . ფორმანტა - აკუსტიკური ენერჯიის კონცენტრაცია მეტყველების ტალღაში კონკრეტული სიხშირის გარშემო.

Abstract

Mathematical methods of speech recognition continue to be an important field with great potential. The increase in computing speed has made it possible to use these methods widely. Spectral analysis is one of the steps in the process of speech recognition, in particular, it means representing the sound signal as a frequency spectrum and analyzing it based on the fast Fourier transform (FFT). Despite the popularity of FFT, from the point of view of practical use, its implementation is associated with large computing resources, especially when complex numbers are involved in the operation. The master thesis compares the efficiency and execution time of FFT and Chebyshev DCT transformation algorithms. In addition, an efficient analysis of vowel and consonant sounds using 256 discrete orthonormal Chebyshev polynomials is proposed. Based on a simple example, it was determined that with the help of 256 DCT, it is possible to convert an analog sound signal from the time domain to the frequency domain, which is a necessary prerequisite in the process of speech recognition. Along with this, using 256 DCT, we can get four formants F_1, F_2, F_3, F_4 at the same time. Formant - the concentration of acoustic energy in a speech wave around a specific frequency.