

# ANALIZA ZVUKA

*Vedran Đerek, 2001./2002. <vex@gkvk.hr>*

## SADRŽAJ

I. Uvod.....	2
II. Teorijski dio – opis fizikalne zakonitosti.....	3
2.1 Osnove mehaničkog valnog gibanja.....	3
2.2 Zvuk.....	5
2.3.1 Fourierove transformacije i spektralna analiza.....	8
2.3.2 Općenito o kvantizaciji vektora i ekstrakciji bitnih osobina zvuka.....	8
III. Eksperimentalni dio.....	11
IV. Zaključak.....	17
V. Popis korištene literature.....	18

# I.

## UVOD

Zvuk, tj. sluh je većini ljudi sasvim svakodnevna pojava. Rijetko kada razmišljamo o važnosti sluha – primanja vanjskih podražaja u obliku zvučnih valova iz sasvim određenog frekvencijskog pojasa. To je jedno od najvažnijih osjetila ljudskog organizma – po važnosti odmah iza vida. Slušne informacije nam omogućavaju lagano sporazumijevanje, orijentaciju i prepoznavanje zvukova.

Razmišljajući o ljudskom sluhu i važnosti sluha i zvukova, kao pravi fizičar, zapitao sam se kako je sve to moguće, i na kojim principima zvuk počiva. Odgovore sam našao, naravno, u obimnoj građi temeljne prirodne znanosti – fizike – koju su pojмили i sročili umovi talentiranih mislilaca-fizičara davno prije moga vremena.

Nakon što sam shvatio principe zvuka, a budući da sam i eksperimentirao sa zvukom na svome računalu, odlučio sam sastaviti kratak i, nadam se, zanimljiv rad o zvuku, kako bih tu zanimljivu pojavu približio onima koji još nisu s njome upoznati, te da bih sistematizirao i ponudio javnosti znanje koje sam prikupio tijekom kratkog razdoblja istraživanja teorije zvuka.

U radu ću se osvrnuti na fizikalne karakteristike zvuka, koristeći se pri tome osobnim računalom sa priključenim mikrofonom i osciloskopom. Tako ću osciloskopom, koji je paralelno priključen sa računalom na mikrofona, pokazati valnu prirodu zvuka, a računalom ću prikazati spektralnu analizu zvučnih valova.

Spektralnu analizu izvodi u stvarnom vremenu računalo putem brzih Fourierovih transformacija (eng. FFT – Fast Fourier Transform), a na ekranu se, također u stvarnom vremenu, iscrtava graf spektralne analize, i očitava se frekvencija zvuka.

Koristeći tu aparaturu objasniti ću boju zvuka, na primjeru nekoliko različitih glazbenih instrumenata. Također, izvesti ću pokus u kojim ću računskim putem predvidjeti broj i frekvencije viših harmonika zvučnog izvora, i taj račun ću potvrditi praktičnim opažanjem pomoću grafa spektralne analize zvuka.

U drugom, zanimljivijem dijelu pokusa prikazati ću jedan vrlo efektivan način primjene analize zvuka – na uzorku od nekoliko (4-5) nasumično izabranih ljudi iz publike pokazati ću kako se spektralna analiza zvuka može iskoristiti za točno prepoznavanje karakteristika glasa svakoga od njih. Računalo će ekstrahirati bitne karakteristike glasa iz uzorka koji će oni dati. Nakon toga, ti "dobrovoljci" će ponovno, nasumičnim redom dati uzorak glasa, i na osnovu toga uzorka i prethodno ekstrahiranih karakteristika, računalo će prepoznati koji govornik je dao uzorak glasa.

Isti pokus se može izvesti i sa nekoliko različitih glazbenih instrumenata – npr. tambura, gitara, flauta...

Vrlo je bitno spomenuti da se ekstrahirane karakteristike glasa ne odnose na riječi koje je govornik izgovorio, već na prave, fizikalne karakteristike tog glasa. Prepoznavanje se ne vrši na osnovi npr. prosječnih vrijednosti frekvencije i amplitude, već na cjelokupnom, statistički reprezentativno odabranom skupu bitnih osobina glasa.

Na samom kraju pokazati ću rezultate mjerenja i nedostatke sistema, i donijeti odgovarajući zaključak.

## II.

# Teorijski dio

## Opis fizikalne zakonitosti

### 2.1

## OSNOVE MEHANIČKOG VALNOG GIBANJA

*Zvuk je val koji se širi prostorom ispunjenim medijem pogodnim za širenje zvučnih valova. Budući da će se u radu koristiti neki pojmovi vezani uz osnove valnog gibanja, u slijedećem poglavlju donosim sažetu valnu teoriju bitnu za daljnje razumijevanje rada.*

Prijenos energije pomoću širenja deformacije u nekom elastičnom sredstvu, naziva se mehaničko valno gibanje. Kod valnog gibanja se kroz sredstvo prenošenja vala giba samo elastični poremećaj, a ne medij – on ostaje nepomičan. Mehanički valovi se mogu širiti kroz elastične medije, u čvrstom, tekućem i plinovitom agregatnom stanju. U vakuumu se mehanički valovi ne mogu širiti, budući da u vakuumu nema medija kroz kojeg bi se valovi širili.

Val je, dakle, gibanje mehaničkog poremećaja kroz elastično sredstvo. Takav val u praksi može biti puls, ili harmonijski val – nas zanima harmonijski val, koji je poremećaj koji nastaje kontinuirano i periodički, iz jednog izvora.

Prostiranje mehaničkog vala može biti dvojako – transverzalno, kada je valni poremećaj okomit na smjer širenja vala, te longitudinalno – valni poremećaj se zbiva u smjeru širenja vala (sl.1).

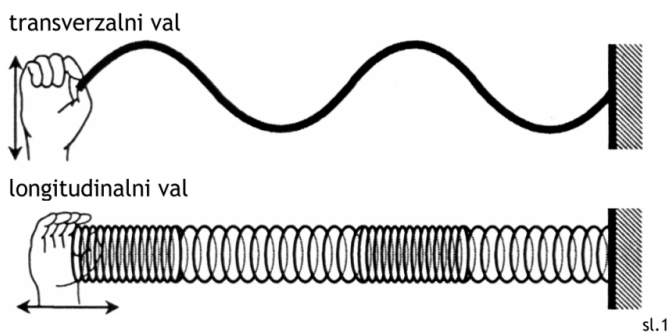
U elastičnim čvrstim sredstvima moguće je širenje i transverzalnih i longitudinalnih

valova, dok se kroz fluide – tekućine i plinove – mogu širiti samo longitudinalni valovi, i to zato što su za postojanje transverzalnog vala nužne sile koje se protive pomicanju jednog sloja prema susjednome. U fluidima su te sile zanemarivo malene ili ne postoje, pa se u njima transverzalni valovi ne mogu razviti.

Kroz zrak, kao fluid, se šire samo longitudinalni valovi. Ti valovi su zvučni valovi.

Svaki harmonijski val možemo smatrati oblikom harmonijskog titranja – zbog toga harmonijski val posjeduje dvije vrlo važne osobine – *vlastitu frekvenciju*  $f$  [Hz] i *valnu duljinu* -  $\lambda$  [m]. Frekvencija harmonijskog vala  $f$  se definira kao broj titraja u sekundi:  $f = \frac{n}{t}$  [Hz], ili kao recipročna vrijednost perioda titranja T:  $f = \frac{1}{T}$  [Hz].

Period titranja T je vrijeme u kojemu se obavi jedan titraj. Valna duljina harmonijskog vala  $\lambda$  [m] je duljina razmaka između dvije točke (čestice) koje titraju sinkrono.

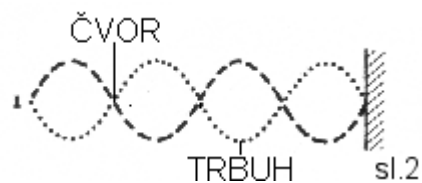


Valovi mogu **interferirati**. To znači da, ako se u prostoru susretnu dva ili više valova, na mjestu presjeka valova nastaje novi val. Taj val odgovara valu što nastaje zbrajanjem elongacija svih pojedinačnih valova što se sijeku u toj točki prostora. Ta se pojava zove interferencija ili superpozicija.

Ako se valovi interferencijom poništavaju, tu interferenciju nazivamo destruktivnom, a ako se valovi interferencijom pojačavaju, nazivamo je konstruktivnom interferencijom.

Poseban slučaj interferencije, kada se u prostoru susretnu dva vala jednakih frekvencija, zove se stojni val.

Glavna karakteristika stojnog vala jest to da se u prikazu amplituda/vrijeme vala pojavljuju specifične točke – trbusi i čvorovi stojnog vala. Čvorovi stojnog vala su točke na kojima je elongacija titranja uvijek jednaka nuli, a točke trbuha vala stalno titraju maksimalnom elongacijom toga vala. Te točke naizgled održavaju stacionaran – stojni položaj. Na slici je primjer stojnog vala koji nastaje na napetoj žici učvršćenoj na oba kraja.



Sustavi koji proizvode stojne valove (npr. napeta žica, stupac zraka) su karakteristični po tome što mogu titrati samo točno određenim frekvencijama. To ograničenje potječe iz same dužine sustava (duljina žice ili stupca zraka). Tako svaki takav sustav može titrati svojom osnovnom frekvencijom – onom na kojoj postoji samo jedan trbuh ili čvor na sredini žice ili stupca zraka (broj čvorova ili trbuha na žici ili stupcu zraka, izuzimajući krajeve žice/stupca, uzima se kao koeficijent  $k$  sustava), ili višim frekvencijama – kada je broj  $k$  veći od jedan. Sve moguće frekvencije titranja sustava koji titra stojnim valovima dane su jednakosti:

$$f_k = \frac{k \cdot v}{2l}$$

gdje je  $k$  broj čvorova ili trbuha na žici ili stupcu zraka (izuzimajući krajnje točke kao trbuhe ili čvorove),  $v$  brzina zvuka u sredstvu, i  $l$  dužina žice ili stupca zraka.

## 2.2 ZVUK

*U ovom poglavlju iznijet ću fizikalne zakonitosti i pojave specifične za zvuk.  
Za medij širenja zvučnih valova uzet ću zrak.*

Zvuk je pojava koju možemo registrirati osjetom sluha – fiziološkim svojstvima uha i interpretacijom slušnih informacija u mozgu. S fizikalnog stajališta, zvuk predstavlja longitudinalni mehanički val u elastičnom sredstvu (zrak), u području frekvencija od prosječno 20Hz do 20000Hz. Valovi frekvencije ispod 20Hz nazivaju se infrazvukom, a oni frekvencije veće od 20kHz ultrazvukom.

Fizikalni (objektivni) parametri zvuka su intenzitet, frekvencija i oblik zvučnog vala. Te objektivne parametre mi osjetilom sluha doživljavamo subjektivno kao glasnoću, visinu i boju (kvalitetu) zvuka.

Intenzitet vala zvuka određen je kao energija koju val pronese kroz jedinicu površine u jedinici vremena:

$$I = \frac{E}{A \cdot t} [W / m^2]$$

Intenzitet zvuka osjet sluha subjektivno doživljava u logaritamskoj skali (Fechnerov zakon). Zato se uvodi pojam razine intenziteta zvuka, koji je definiran kao deseterostruki logaritam omjera intenziteta zvuka i intenziteta zvuka na pragu čujnosti:

$$L = 10 \log \frac{I}{I_0} [dB]$$

Subjektivno doživljena visina tona izravno je ovisna o frekvenciji zvučnog vala, ali vrlo rijetki ljudi posjeduju mogućnost točnog određivanja frekvencije zvuka – to je tzv. "apsolutni sluh".

Oblik zvučnog vala u praksi nije savršeno sinusoidalan, već je više ili manje nepravilnog oblika – uzorka koji se neprestano ponavlja. Za istu frekvenciju različiti glazbeni instrumenti, kao i ljudski glasovi, daju različit oblik vala. Taj različit oblik vala doživljavamo kao boju tona, pomoću koje raspoznavamo različite glasove, glazbene instrumente itd. Ta nepravilnost u obliku vala kod npr. napetih glasnih žica u govornom aparatu čovjeka ili kod napetih žica gitare, proizlazi iz toga da napeta žica istovremeno titra na nekoliko različitih frekvencija – stvara se stojni val složena oblika, koji je nastao kao rezultat interferencije više valova koji titraju osnovnom i višim frekvencijama. Podsjetimo se, te frekvencije su uvijek jednake cjelobrojnom umnošku osnovne frekvencije. Osnovna frekvencija ima najveću amplitudu, a amplituda nižih harmonika obično pravilno opada s povećanjem frekvencije, ali to nije nužno, poglavito zbog osobite građe rezonantnih struktura koje su povezane sa izvorom titranja (drvena kutija gitare, usna šupljina i grlo kod čovjeka).

No, te nepravilnosti u raspodjeli viših harmonika mogu biti vrlo korisne. To ću objasniti u slijedećem poglavlju.

## 2.3

### 2.3.1 FOURIEROVE TRANSFORMACIJE I SPEKTRALNA ANALIZA

Spektralna analiza zvuka je postupak kojim se zvučni val razlaže na njegove sastavne frekvencije. Ulazni parametri spektralne analize su amplituda zvuka u vremenu, a izlazni – raspodjela frekvencija u određenom frekventnom pojasu u vremenu.

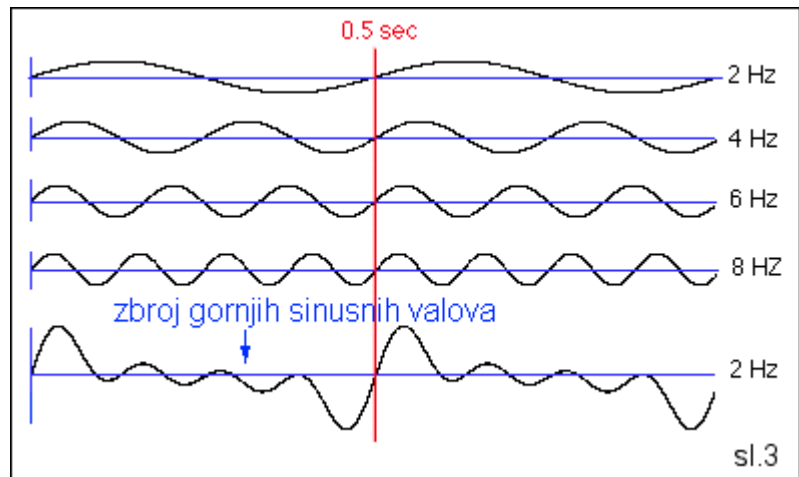
Spektralnu analizu je najjednostavnije shvatiti na primjeru stojnog vala koji nastaje npr. kod napete žice na glazbenom instrumentu. Kao što sam već spomenuo, žica na instrumentu u praksi ne titra samo jednom frekvencijom, već sa više mogućih frekvencija, od kojih najjaču amplitudu ima ton osnovne frekvencije, a nakon njega se nižu viši harmonici sa postupno padajućim amplitudama. Kako tu spoznaju iskoristiti za spektralnu analizu zvuka?

Pogledajmo sliku 3:

Na slici se vide pojedinačni sinusni valovi, frekvencija 2,4,6 i 8 Hz. Gornji val od 2 Hz predstavlja osnovni ton, a valovi od 2,4,6 i 8 Hz predstavljaju više harmonike toga osnovnog tona.

Rezultat zbrajanja tih valova je donji val, frekvencije 2 Hz. Taj val je, po svom izgledu, vrlo sličan nepravilnim valovima koji

nastaju npr. govorom ili na glazbenim instrumentima. Bitna značajka tog vala je da on u sebi, iako je naizgled nepravilan, sadržava informaciju od kojih je harmonika sastavljen, kao i njihovu amplitudu. Zbrajanje pojedinih valova u rezultatni val u praksi je interferencija tih valova. Znači, neka funkcija inverzna toj funkciji bi dala raspodjelu i amplitudu osnovnog tona i viših harmonika – vršila bi spektralnu analizu. Bitno je primijetiti da je domena složenog vala kao funkcije vrijeme, a domena funkcije inverzne funkciji interferencije kojom nastaje taj složeni val – frekvencija.



Prirodu te inverzne funkcije je objasnio barun *Jean Baptiste Joseph Fourier* (1768-1830) (sl.4). Naime, on je u svome djelu "*Théorie analytique de la chaleur*" (*Analitička teorija topline*) objavljenom 1822. godine zaključio da se bilo koja periodična funkcija  $y=f(x)$  perioda  $\lambda$  može aproksimirati trigonometrijskim polinomom – tzv. **FOURIEROVIM REDOM**. Taj postupak se naziva Fourierovom ili harmonijskom (spektralnom) analizom.

Ako za periodičnu funkciju uzmemo zvučni val, tada se, po Fourieru, svaki rezultatni zvučni val može aproksimirati trigonometrijskim redom oblika:

$$y = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx$$

gdje su  $a_0$ ,  $a_n$  i  $b_n$  koeficijenti Fourierove konstante. Ti koeficijenti su integrali racionalnih funkcija, i njihovo glavno svojstvo je da pri računanju daju najmanju pogrešku. Izrazi za te koeficijente se izvode integralnim računom, a budući da ti izrazi nisu neophodni za razumijevanje spektralne analize, neću ih niti navoditi.

No, ovim osnovnim oblikom Fourierova reda nije se praktično koristiti, pogotovo ne za izračunavanje spektralne analize složenih zvučnih valova, i to iz dva razloga – on sadrži nezgrapne koeficijente, koji su i sami funkcije – njih bi bilo vrlo nezgodno računati, a također, računanje bi trajalo beskonačno. A čak da moderna računala i podržavaju rad s beskonačnostima, sumnjam da bismo imali toliko vremena za izvođenje računa.

Upravo zato se za računanje spektralne analize na računalima najčešće koristi transformirana verzija gornje jednačbe koja ima ograničen broj parametara:

$$y_p = \sum_{k=0}^{n-1} x_k \left[ \cos\left(2\pi \frac{kp}{n}\right) + i \sin\left(2\pi \frac{kp}{n}\right) \right]$$

gdje je  $x_k$  k-ta kompleksna vrijednost unosa,  $y_p$  p-ta kompleksna vrijednost rezultata funkcije i  $n = 2^N$  ukupan broj uzoraka na kojima se obavlja račun. Taj oblik Fourierovog reda se dobiva prilagođavanjem i deriviranjem originalne jednačbe. No, postupak izvođenja je prilično složen, pa donosim samo rezultat transformacija. Taj algoritam se naziva BRZA FOURIEROVA TRANSFORMACIJA (FFT), upravo zato što je nastao kao rezultat transformiranja Fourierova reda. Tim algoritmom se postižu najbrža izračunavanja, ali on ima i neka značajna ograničenja – npr. broj uzoraka na kojemu se obavlja račun nužno mora biti jednak  $2^N$ , gdje je N cijeli broj.

Rezultat brze Fourierove transformacije je skup kompleksnih brojeva. Da bi se iz takvog prikaza prešlo u prikaz pogodan za prikaz spektralne analizu, upotrebljava se tzv. *power spectrum* algoritam. Pri konverziji u oblik u kojemu je domena frekvencija, power spectrum algoritam predstavlja frekvenciju kao redni broj kompleksnog broja u skupu, a elongaciju kao zbroj kvadrata realnog i imaginarnog dijela kompleksnog broja.

No, kod praktične primjene ove metode javlja se problem – naime, zvučni valovi nisu niti približno periodični, pogotovo ne na razini malih uzoraka od nekoliko milisekundi. Da bi se postigla virtualna periodičnost uzorka, te samim time poboljšala preciznost računa, na uzorak se primjenjuje metoda *windowing funkcija* (eng.) – te funkcije djeluju tako da amplitude krajnjih dijelova uzorka smanje, tako da se one, idući prema sredini uzorka, postupno povećavaju ili smanjuju. Kada se takav uzorak provede kroz FFT funkciju, ona ga može predstaviti kao periodični signal ako

uzmemo da je period tog uzorka jednak njegovoj duljini. Također, windowing funkcije u praksi sprječavaju "curenje" signala u susjedne spektralne linije, ali njihova loša strana je u tome što smanjuju razlučivost analize. No, postoje windowing funkcije koje zadovoljavaju oba kriterija. U ovom radu koristi se optimizirana Blackman windowing funkcija:

$$y_n = \sum_{n=0}^{n=N} x_n + \frac{\alpha+1}{2} - \frac{1}{2} \cos\left(\frac{2\pi n}{N-1}\right) - \frac{\alpha}{2} \cos\left(\frac{4\pi n}{N-1}\right)$$

gdje je  $y_n$  rezultatna amplituda n-tog člana skupa,  $x_n$  amplituda n-tog člana skupa, N veličina uzorka i  $\alpha$  konstanta funkcije – optimalna konstanta iznosi:

$$\alpha = - \left[ \frac{\sin \frac{\pi}{N-1}}{\sin \frac{2\pi}{N-1}} \right]^2$$

Maksimalan frekvencijski opseg spektralne analize osnovane na FFT metodi je jednak polovici frekvencije uzorkovanja (sampling rate). Ako se uzme frekvencija uzorkovanja potrebna za CD kvalitetu zvuka – 44100 Hz, frekvencijski opseg je od 0-22050 Hz, što je u osnovi maksimalan opseg ljudskog sluha.

Preciznost FFT spektralne analize – njena razlučivost – je jednaka kvocijentu broja članova u uzorku i frekvencije uzorkovanja. To proizlazi iz toga da je fundamentalna frekvencija FFT analize, znači i njena razlučivost, obrnuto proporcionalna periodu koji predstavlja uzorak unesen u FFT algoritam. Taj period je jednak kvocijentu broja članova u uzorku i frekvencije uzorkovanja, pa je, stoga, fundamentalna frekvencija i razlučivost analize obrnuto proporcionalna toj vrijednosti.

## 2.3.2 OPĆENITO O KVANTIZACIJI VEKTORA I EKSTRAKCIJI BITNIH OSOBINA ZVUKA

U drugom dijelu pokusa pokazat ću moguću primjenu spektralne analize zvuka pri postupku glasovne identifikacije. Praktična upotreba glasovne identifikacije može biti u projektima vezanim uz sigurnost, nadgledanje, a ponekad se koristi i u sudskoj praksi.

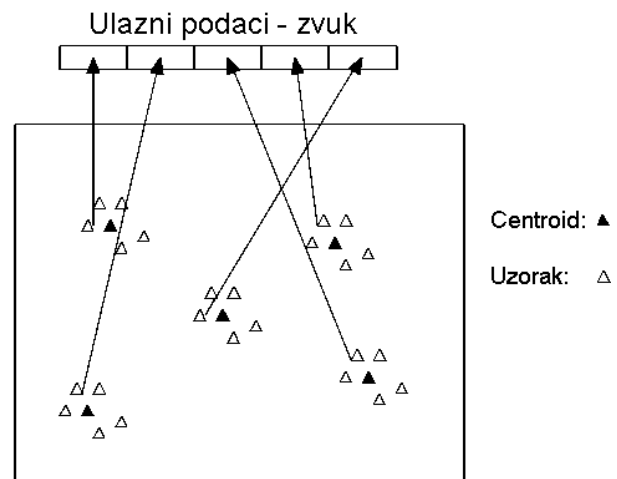
Proces glasovne identifikacije je vrlo složen proces, i postoji više pristupa u njegovom rješavanju. U ovom projektu se koristi pristup sa kvantizacijom vektora, zbog visoke preciznosti i lake implementacije.

Da bi bila moguća glasovna identifikacija, potrebno je proći kroz nekoliko koraka – snimanje glasa i predstavljanje govornika sistemu, zatim ekstraktiranje bitnih osobina glasa i njihovo spremanje u bazu podataka kao prvi korak, te ponovno snimanje glasa i ekstraktiranje bitnih karakteristika, koje se onda uspoređuju sa onima u bazi podataka, te se utvrđuje identitet.

Ova dva koraka se u biti sastoje od tri bitna procesa: snimanja glasa, ekstraktiranja bitnih osobina iz dobivene snimke, koje se odvija u dva dijela, i, na kraju, uspoređivanja glasova na temelju sličnosti, na temelju čega se izvodi zaključak o identitetu.

Snimanje glasa se obavlja pomoću mikrofona spojenog na računalo. Snimanje se vrši u 16 bitnom načinu, uz frekvenciju uzorkovanja od 16000 Hz. Pri predstavljanju govornika računalo snima se desetak sekundi glasa, da bi se dobio reprezentativni i ujednačen uzorak za daljnju obradu, a pri provjeri identiteta snima se samo 4 sekunde govora, jer je sasvim mala količina podataka potrebna za daljnji postupak.

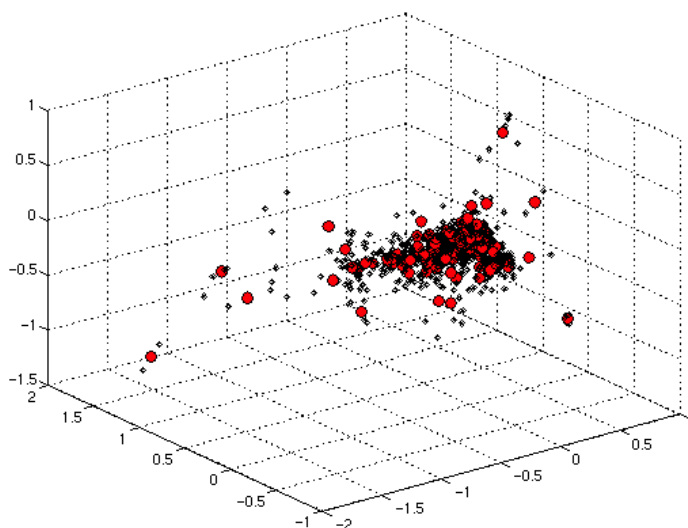
Ekstrakcija bitnih osobina glasa je zapravo svođenje uzorka na manju veličinu ali uz očuvanje dobre reprezentativnosti uzorka. To se izvodi u dva dijela – u prvom dijelu se snimljeni glas provodi kroz algoritam brze Fourierove transformacije, te se dodatnim frekvencijskim filtriranjem uklanja pozadinska buka i manje nepravilnosti i varijacije glasa, i kao rezultat dobivaju se tzv. *koeficijenti govornog aparata* – eng. *cepstral coefficients* – podaci (brojevi) koji vjerno predstavljaju osobine govornog aparata govornika. Poseban značaj ovog postupka je u tome što bitne osobine glasa ostaju sačuvane, a znatno se smanjuje količina informacija koja ih predstavlja – ovim postupkom se zapravo odbacuju nebitne informacije i smanjuje količina podataka na kojoj će kasnije biti vršeni složeni i vremenski zahtjevni proračuni.



sl.5

Drugi dio obrade glasovnog uzorka obuhvaća postupak tzv. *kvantizacije vektora* (eng. *vector quantization*). Pojam vektora ovdje ne znači klasičan, matematički vektor, nego ulaznu informaciju. Kvantizacija vektora služi još većem smanjivanju i povećavanju informacija dobivenih primarnom ekstrakcijom.

Kompresija informacija je drastična, ali su bitne informacije o prirodi zvuka još uvijek vrlo vjerno prikazane. Ovu metodu je moguće objasniti na priloženoj slici (sl. 5) – iz podataka dobivenih u prvom koraku dobiva se nepravilno raspoređena struktura (bijeli trokutići). Položaj najveće gustoće tih trokutića – uzoraka – vjerno prikazuje jedan tzv. *centroid* vektor. Na taj način vjernost informacija ostaje visoka, a količina podataka potrebnih za njihovo prikazivanje se drastično smanjuje. To zorno prikazuje i slika 6 – veći kružići predstavljaju centroid vektore, a manji kružići predstavljaju ulazne podatke – uzorke.



sl.6

Za izvođenje računa kvantizacije vektora koristi se nekoliko metoda – *binary splitting* (na principu dijeljenja uzorka u polovine koje se optimiziraju, sve dok se ne postigne željeni stupanj preciznosti), *random coding* (nasumičan odabir centroid vektora sve dok se ne postigne željeni rezultat) i trenutno najbolje izveden algoritam, tzv. *LBG* (Linde-Buzo-Gray) algoritam, poznat i kao *optimization with k-means* i *generalizirani Lloydov algoritam*. Taj algoritam se koristi u ovome radu. Objašnjenje tog algoritma je prilično opširno, i opsegom ne spada u ovaj rad.

Nakon ekstrakcije bitnih osobina glasa, dobivamo rezultate u obliku tzv. *codebooka* ili *modela*, koji predstavljaju vrlo sažetu informaciju o osobinama zvuka. Postupkom ekstrakcije na sistemu koji sam ja implementirao postiže se stupanj sažimanja, tj. reduciranja nepotrebnih informacija, od 85%. Ta brojka se po potrebi može još više povećavati bez značajnijeg utjecaja na preciznost sistema.

Takvim modelima se može uspoređivati sličnost mjerenjem akumulirane udaljenosti specifičnih centroid vektora – što je udaljenost manja, sličnost je veća. Na tom principu počiva sam postupak glasovne identifikacije.

Pri pisanju programa za identifikaciju glasa, naravno, nije bilo niti moguće, a niti potrebno, da sam pišem kod za implementaciju npr. FFT-a, MFCC algoritma ili LBG algoritma. To, uostalom, nije ni znanstvena, ni programerska praksa.

Pri pisanju programa poslužio sam se Intelovom® bibliotekom funkcija za obradu signala (Intel® Signal Processing Library), zbog brzine izvršavanja i prilagođenosti Intelovim® procesorima. Tu biblioteku funkcija koristio sam za izračunavanje brzih Fourierovih transformacija u stvarnom vremenu. Za identifikaciju glasa koristio sam biblioteku funkcija SVLib, "Speaker Verification Library", čiji mi je izvorni kod na uvid ljubazno ustupio autor te biblioteke funkcija, znanstvenik Jialong He sa "Arizona State University" sveučilišta u Arizoni, SAD. Iz nje sam koristio algoritme za ekstrakciju osobina glasa (MFCC - *Mel-scaled FFT based*

cepstrum algoritam) te algoritam za kvantizaciju vektora (posebno LBG algoritam). Te dvije biblioteke funkcija sam uključio u svoj program, koji obavlja snimanje te kontrolira i koordinira izvođenje operacija potrebnih za glasovnu identifikaciju, koristeći funkcije iz biblioteka funkcija, te sprema gotove modele (*codebooke*) na *hard disk* računala.

Iako su sve bitne funkcije koje sam koristio sadržane u bibliotekama funkcija, njihova upotreba jednostavno nije moguća bez dobrog poznavanja načina njihova rada.

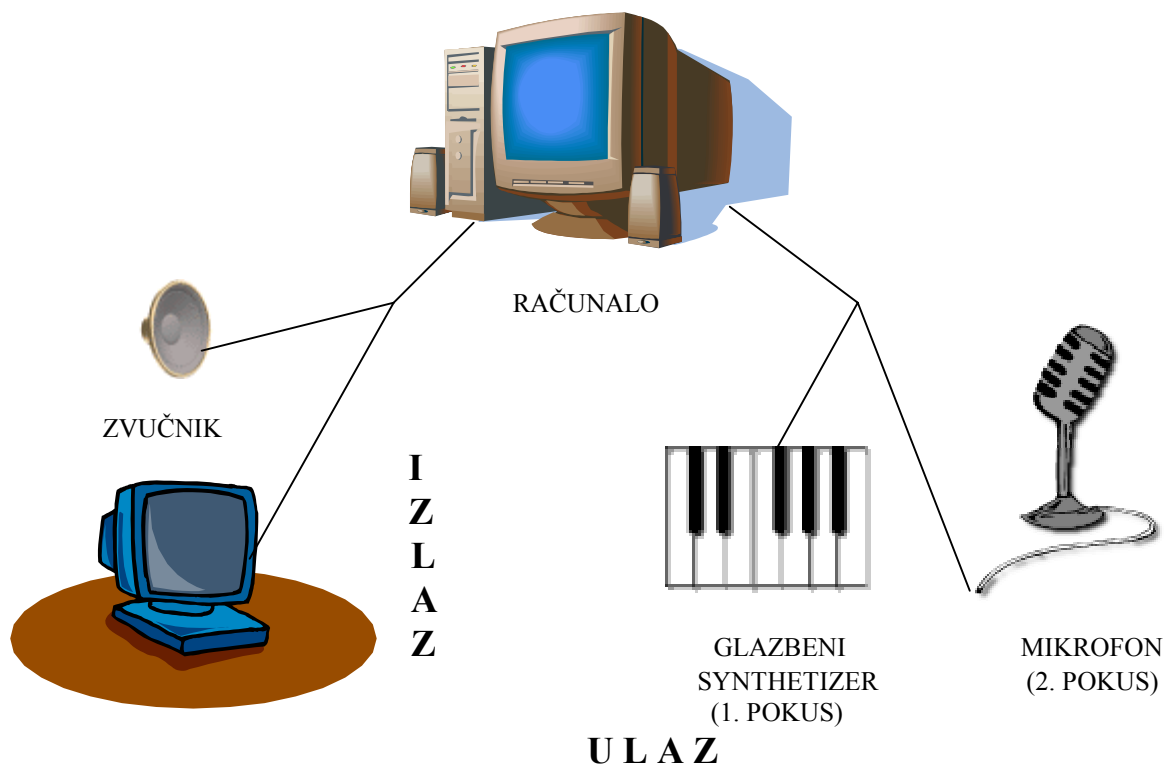
Program, kao i biblioteke funkcija, je napisan programskim jezikom "C" i "C++".

### III.

## Eksperimentalni dio

- Pribor: - mikrofons  
- glazbeni sintetizer  
- osciloskop  
- osobno računalo (PC PII) spojeno na TV-ekran, sa pripadajućim softwareom

### SHEMA POKUSA



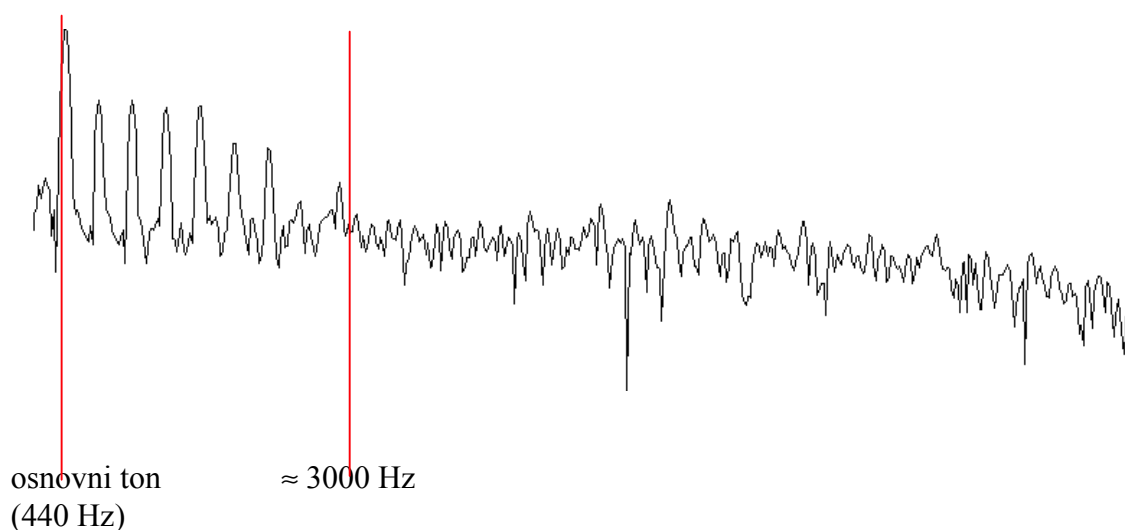
- TV-EKRAN (SPEKTRALNA ANALIZA)  
I OSCILOSKOP (1. POKUS)  
- PROGRAM ZA ANALIZU GLASA (2. POKUS)

## **POKUS BR. 1 (uvodni pokus)**

**OPIS POKUSA:** Na glazbenom synthesizeru se redom za pet različitih instrumenata (glasovir, trubu, triangl i crkvene orgulje), odsvira nota A (440 Hz). Primjećuje se različit izgled oblika zvučnog vala na osciloskopu. Računalom se istovremeno izvodi spektralna analiza putem brzih Fourierovih transformacija (FFT veličine 2048 uzoraka, 44100 Hz frekvencija uzorkovanja, frekvencijski opseg 0-22050 Hz). Valja primijetiti da je za sve instrumente na grafu spektralne analize osnovna frekvencija ista – 440 Hz. Primjećuje se i pravilan raspored viših harmonika. Mjeri se razlika frekvencija viših harmonika i donosi se zaključak.

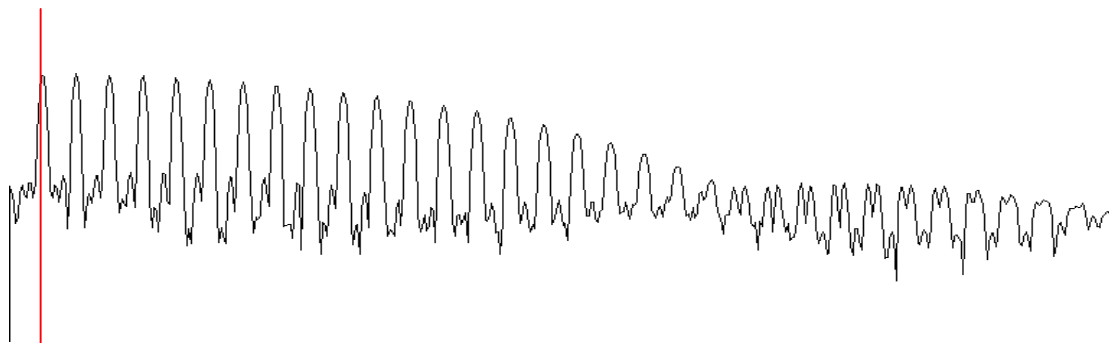
### **REZULTATI MJERENJA:**

#### **- GLASOVIR:**



- Izmjerena prosječna osnovna frekvencija (3 mjerenja): 441 Hz
- Mjerenja ukazuju na to da se formira pravilan stojni val osnovne frekvencije  $\approx 440$  Hz, te viši harmonici te frekvencije. Na grafikonu se raspoznaje 8 viših harmonika uz osnovnu frekvenciju. Amplituda opada strmo od osnovnog tona (grafikon je prikazan u logaritamskoj skali), i opada prema višim harmonicima. Greška između mjerenih (3 mjerenja) i očekivanih rezultata iznosi 13%. Nema uočljivih vrlo visokih harmonika (iznad 3000 Hz).

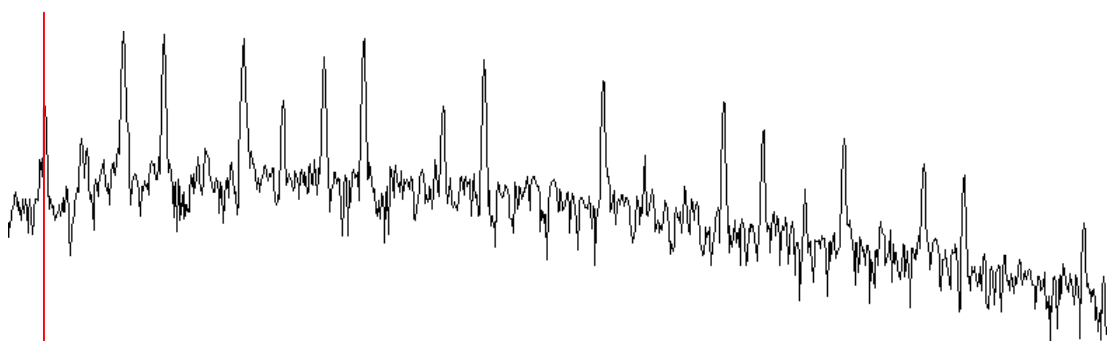
- TRUBA: osnovni ton (440Hz)



osnovni ton (440 Hz)

- Izmjerena osnovna frekvencija je identična onoj izmjerenoj za glasovir (441.5±9.5 Hz)
- Ponovno se pokazuje da nastaje stojni val osnovne frekvencije 440 Hz sa pripadajućim višim harmonicima, koji su pravilno raspoređeni duž cijelog spektra. Primjećuje se znatno smanjenje amplitude oko 7000 Hz, a nakon te frekvencije ponovno se pojavljuju ponešto distortirani viši harmonici.
- Greška između očekivanih i izmjerenih rezultata iznosi 12%

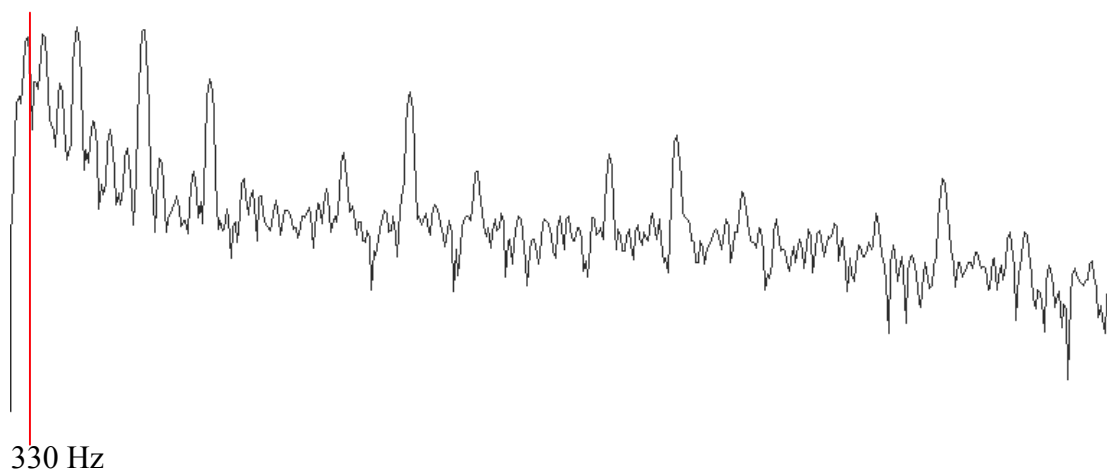
- TRIANGL:



900 Hz

- Izmjeren je osnovni ton od 900 Hz
- Viši harmonici se pojavljuju u pravilnim frekvencijskim razmacima, ali sa vrlo nepredvidivim amplitudama. Vidljive su serije amplituda po višim harmonicima koje se ponavljaju – velika – mala – dvije velike – mala amplituda.

- CRKVENE ORGULJE:



- Jedna od osnovnih frekvencija je 330 Hz, druga je 440 Hz
- Zbog prirode građe orgulja (više svirala zvuče u isto vrijeme) vidi se nekoliko osnovnih tonova sa svojim pripadajućim višim harmonicima. Primjećuje se i pravilo u pojavi većih amplituda – grupirane su po dvije u pravilnim razmacima. To dovodi do zaključka da ton A na ovim orguljama nastaje istovremenim zvučanjem barem dvije veće cijevi (svirale) na orguljama.

**ZAKLJUČAK:** Mogli smo utvrditi istinitost valne teorije zvuka, i pokazati funkcionalnost sustava. Primjećuje se karakterističan izgled vala za svaki glazbeni instrument, kao i različiti grafovi spektralne analize. Na temelju tih podataka moguće je vršiti analizu pojedinih zvukova i njihovih karakteristika – npr. duljine i napetosti žice, duljine svirale, broj žica/svirala koje zvuče istovremeno itd. Preciznost sustava – njegova razlučivost – je približno 21 Hz, što znači da sustav mjeri frekvencije u koracima od 21 Hz. Iz toga proizlaze odstupanja u točnosti između izmjerenih i izračunatih rezultata. Ipak, greška nije velika ako uzmemo u obzir veliki frekvencijski raspon od 22050 Hz.

## POKUS BR.2

**OPIS POKUSA:** Za ovaj pokus potrebno je četiri osobe–statista iz gledališta (komisije). Svaka od osoba mora dati uzorak svoga glasa u trajanju od sedam sekundi. Nije bitno što osoba govori, već kako govori – bitno je održati normalan ton glasa. Tom uzorku glasa se u bazi podataka na računalu pridružuje ime te osobe. Taj postupak se izvede za tri osobe.

Nakon toga, sve četiri osobe ponovno daju svoj uzorak glasa, ovaj puta u trajanju od četiri sekunde, nasumičnim redoslijedom, da bi se izbjegla mogućnost manipulacije rezultatima. Računalo na osnovi glasovnih informacija sadržanih u drugom uzorku koje su osobe dale, i na osnovi već sačuvanih informacija u bazi podataka, identificira svakog od govornika, osim onoga koji nije na početku dao uzorak. Taj četvrti govornik služi kao kontrolni faktor, i njega sustav identificira kao NEPREDSTAVLJENOG GOVORNIKA.

**REZULTATI MJERENJA:** Opsežnije mjerenje je izvedeno na uzorku od 25 osoba. Na osnovi tog mjerenja može se zaključiti da sustav glasovne identifikacije ima preciznost od prosječno 83%. Gubitak preciznosti je posebno izražen kod loše kvalitete snimke glasa, kod vrlo tihog govora, i onda kada je veličina uzorka glasa malena. No, kod valjanih i kvalitetnih ulaznih informacija preciznost je vrlo velika. Sustav ima određenih problema sa većim brojem govornika u bazi podataka, jer se vrijeme potrebno za identifikaciju povećava proporcionalno broju govornika u bazi podataka (vrijeme potrebno za identifikaciju jednog govornika je oko 6 sekundi). U slijedećoj tablici donosim reprezentativni izvadak iz mjerenja (5 različitih govornika, svaki govori po 2 rečenice od 4 sekunde):

GOVORNIK/REČENICA (veći broj znači veću istinitost identifikacije)										
GOVORNIK	1 / 1	1 / 2	2 / 1	2 / 2	3 / 1	3 / 2	4 / 1	4 / 2	5 / 1	5 / 2
1	<b>19.04</b>	<b>18.91</b>	17.10	15.96	16.10	15.27	17.16	17.21	17.56	18.67
2	17.83	17.94	<b>18.12</b>	18.11	17.24	17.02	17.15	17.35	17.50	<b>18.20</b>
3	16.33	17.72	15.78	16.46	<b>18.40</b>	<b>19.25</b>	15.89	16.34	18.02	16.72
4	17.98	18.2	17.59	17.10	17.22	16.54	18.09	<b>18.60</b>	18.14	18.65
5	18.02	17.6	17.00	16.20	16.60	16.70	16.20	16.17	<b>18.89</b>	<b>19.76</b>

TUMAČ: Svaki redak (od 1-5) predstavlja posebnog govornika iz baze podataka  
Stupci predstavljaju redni broj testne rečenice govornika (1/1 znači: 1. govornik iz baze podataka, 1. testna rečenica).  
Koso iscrtana polja označavaju realno podudaranje govornika, a tamno obojena identificiranog govornika.  
Koso iscrtana polja sa masno otisnutim brojkama označavaju točnu identifikaciju. Neobojana polja označuju negativnu identifikaciju.  
Vrijednosti u tablici predstavljaju recipročnu vrijednost akumulirane udaljenosti pri postupku kvantizacije vektora između govornika u bazi podataka i testiranog govornika. Veći broj znači veću sličnost između govornika u bazi i testiranog govornika.

Iz tablice se može izračunati prosječna točnost sustava (8 točnih od 10 mogućih identifikacija – približno 80%, što odgovara rezultatu mjerenja na većem broju ispitanika – 83%). Primjećuje se i relativno mala razlika između točno identificiranog govornika i govornika najbližeg njemu. To objašnjavam prilično lošom kvalitetom snimke (govor je sniman u prostoriji (učionici) sa dosta šuma, mikrofonom niske kvalitete, neki govornici su namjerno željeli prevariti sustav). U idealnim uvjetima (npr. snimanje u tonskom studiju na profesionalnoj opremi) i rezultati bi bili idealni. No, čak i u ovako lošim uvjetima sustav je pokazao veliku razinu pouzdanosti.

NAPOMENA: Bitno je reći da se ovakav sustav može koristiti u dvije namjene – za identifikaciju govornika, što sam pokazao, te za verifikaciju, tj. potvrđivanje identiteta govornika. Verifikacija je postupak potvrde identiteta – pri verifikaciji računalu je poznat navodni identitet osobe koja se predstavlja (npr. identifikacijski broj sa slip-kartice), i računalu samo treba potvrditi da li je govornik onaj za koga se predstavlja. Pokusi sa takvim sustavom su pokazali veliku točnost takvog sustava, ali njegova točnost jako ovisi o kvaliteti, te o konstantnosti kvalitete ulaznih informacija. Na tom polju je potrebno dodatno usavršavanje, koje se može ostvariti, što potvrđuju i brojne stvarne primjene ovakvih sustava.

#### IV.

## Zaključak

Na primjerima pokusa potvrđuje se važnost i gotovo neograničena mogućnost praktične primjene spektralne analize zvuka u mnogim granama suvremene znanosti, industrije i tehnologije. Spektralna analiza nalazi svoju veliku primjenu i u građevini, obradi glazbe, sudskoj medicini, kao i u svim drugim primjenama u kojima je potrebno utvrditi porijeklo zvuka. Postupak spektralne analize prošao je dug i trnovit put od dana kada je čuveni znanstvenik – matematičar i fizičar Jean Baptiste Joseph Fourier utvrdio principe spektralne analize, koji još dugo nisu bili prihvaćeni, do primitivnih metoda mehaničkih sprava i tablica za ručno računanje spektra. No, u zadnjih 30-40 godina, intenzivnim razvojem informatičke industrije i eksponencijalnim povećavanjem procesne snage računala, intenzivno se razvija i metoda spektralne analize, zajedno sa novim, brzim algoritmima. Moderna računala, kao što je ovo uporabljeno u pokusu, izvode stotine milijuna računskih operacija u sekundi. Tek takvom procesnom snagom je moguće izvesti tako složeni račun, kao što je računanje spektralne analize, u realnom vremenu. Također, tek sada su razvijeni pravi sustavi za glasovnu identifikaciju, i polako nalaze najrazličitije primjene – a njihovo vrijeme tek dolazi. Vjerujem da će vrlo skoro proizvođači ključeva i brava ustupiti svoje mjesto proizvođačima softvera za glasovnu identifikaciju.

## Popis korištene literature

- [1] Dr. Željko Marković, "Uvod u višu analizu" I dio, 4. izdanje, Školska knjiga, Zagreb, 1956.
- [2] Mladen Paić, "Predavanja iz mehanike II", Sveučilište u Zagrebu, Zagreb, 1969.g.
- [3] Prof. dr. Žarko Dadić, "Povijest ideja i metoda u matematici i fizici", Školska knjiga, Zagreb, 1992.
- [4] Dr. Rudolf Krsnik – Branka Mikulčić, "Međudjelovanja, relativnost, titranja i zvuk", Školska knjiga, Zagreb, 1992.
- [5] Mikulčić – Varićak – Vernić, "Zbirka zadataka iz fizike", Školska knjiga, Zagreb, 1996.g
- [6] akad. Vladimir Paar, "Fizika 3", Školska knjiga, Zagreb, 1998.
- [7] Kljajić – Kolesarić – Krizmanić – Szabo – Zarevski – Žužul, "Psihologija", VII. izdanje, Školska knjiga, Zagreb, 1998.
- [8] Boettcher – Soong – Siegendorf, "Text-Independent Speaker Identification", Dept. of Electrical and Computer Engineering at Carnegie Mellon University, Pittsburgh, PA, 1998.
- [9] Jialong He, "Speaker verification library" ([http://www.bigfoot.com/~Jialong\\_He](http://www.bigfoot.com/~Jialong_He)), 2000. (dokumentacija i izvorni kod)
- [10] Intel ® Signal Processing Library *Reference Manual*, Copyright © 1995-1999 Intel Corporation, Issued in U.S.A., Order Number 630508-010, web: <http://developer.intel.com>
- [11] Linde Y., Buzo A., and Gray R. M. "An algorithm for vector quantizer design," IEEE Trans. Comm., Vol. 20, pp. 84-95, 1980.
- [12] J. Picone, "Signal modeling techniques in speech recognition", Proceedings of the IEEE, Vol. 79, No. 4, April 1991.
- [13] Van Loan, C. 1992, "Computational Frameworks for the Fast Fourier Transform" (Philadelphia:S.I.A.M.).