SVEUČILIŠTE U ZAGREBU FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 1645

### OTKRIVANJE I PRAĆENJE GOVORNIKA U OKOLINI MOBILNOG ROBOTA POLJEM MIKROFONA

Ivan Marković

Zagreb, rujan 2008.

Sveučilište u Zagrebu Fakultet elektrotehnike i računarstva Zavod za automatiku i računalno inženjerstvo

Diplomski rad br. 1645

# Otkrivanje i praćenje govornika u okolini mobilnog robota poljem mikrofona

# Ivan Marković

Mentor: Prof. dr. sc. Ivan Petrović

Zagreb, 2008.

... 😳

# Sažetak

U ovom diplomskom radu predložena je metoda lokalizacije i praćenja govornika, koja se temelji na estimaciji razlike u vremenu dolaska signala (engl. *Time Difference of Arrival*, TDOA). Prvo, TDOA estimacija je poboljšana s tzv. β-faznom transformacijom (engl. *Phase Transform*, PHAT) te testirano s dvoelementnim poljem mikrofona. Metoda lokalizacije se temelji na pseudo-linearnom estimatoru (engl. *Pseudo-Linear Estimator*, PLE), također je predloženo polje Y geometrije za prostorno uzorkovanje, te je uspoređeno s kvadratnim poljem mikrofona. Praćenje je ostvareno s algoritmom rekurzivnih najmanjih kvadrata (engl. *Recursive Least Squares*, RLS). Na kraju, rezultati dobiveni snimanjem na mobilnom robotu su predstavljeni, pokazujući da je razvijeno audio sučelje i algoritam sposobno lokalizirati i pratiti govornika u okolini mobilnog robota.

**Ključne riječi:** estimacija azimuta, TDOA estimacija, DOA estimacija, hiperbolička multilateracija, otkrivanje govornika, polje mikrofona.

# Sadržaj

Sažetak								v		
Sa	ıdrža	aj								i
1	Uvo	od								3
	1.1	Motivacija i cilj								3
	1.2	Pregled rada			•	•	• •	•		5
2	Teorija lokalizacije							7		
	2.1	Prostorno-vremenski signali								7
	2.2	Zvučno polje kao prostorno-vremenski sigr	nal							8
	2.3	Šum i odjek			•	•	• •	•	•	11
3	TDOA estimacija					15				
	3.1	TD estimacija								15
	3.2	Metode određivanja TDOA								16
	3.3	Generalizirana kroskorelacija (GCC)								18
		3.3.1 Otežavanje funkcije kroskorelacije								19
		3.3.2 PHAT- $\beta$ težinska funkcija								20
	3.4	Pregled algoritma za određivanje TDOA .				•		•	•	21
4	DOA estimacija							23		
	4.1	Estimacija smjera dolaska								23
	4.2	Detektor govorne aktivnosti								25
	4.3	Rezultati eksperimenta DOA estimacije								25
	4.4	Test 0°								28
		4.4.1 Pucketanje prstima				•				28
		4.4.2 Izgovor "Test"								29

		4.4.3 Izgovor "Test jedan, dva, tri"	30
	4.5	Test -15°	32
		4.5.1 Pucketanje prstima	32
		4.5.2 Izgovor "Test"	33
		4.5.3 Izgovor "Test jedan, dva, tri"	34
	4.6	Test -30°	36
		4.6.1 Pucketanje prstima	36
		4.6.2 Izgovor "Test"	37
		4.6.3 Izgovor "Test jedan, dva, tri"	38
	4.7	Test -45°	40
		4.7.1 Pucketanje prstima	40
		4.7.2 Izgovor "Test"	41
		4.7.3 Izgovor "Test jedan, dva, tri"	42
	4.8	Test -60°	44
		4.8.1 Pucketanje prstima	44
		4.8.2 Izgovor "Test"	45
		4.8.3 Izgovor "Test jedan, dva, tri"	46
	4.9	Test 70°	48
		4.9.1 Pucketanje prstima	48
		4.9.2 Izgovor "Test"	49
		4.9.3 Izgovor "Test jedan, dva, tri"	50
	4.10	Test -80°	52
		4.10.1 Pucketanje prstima	52
		4.10.2 Izgovor "Test"	53
		4.10.3 Izgovor "Test jedan, dva, tri"	54
	4.11	Analiza rezultata eksperimenata	56
5	Hip	erbolička multilateracija 5	57
	5.1	Hiperbola kao položaj govornika	57
	5.2	Hiperbolička estimacija pozicije	59
	5.3	Lokalizacija govornika metodom sjecišta hiperbola 6	60
<i>с</i>	A 1		
0	Algo	oritam lokalizacije i pracenja govornika	b5
	6.1	Lokalizacija govornika metodom sjecista asimptota niperbola	60
	0.Z	Diskrimina cija actimacija azimuta	00 60
	0.3	Diskriminacija estimacije azimuta	0ソ 71
	0.4 6 E	Algoritam preferie gevernike	/ 1 72
	0.3	Estreorimentalni regultati kensărea ale aritme	13
	6.6	Eskperimentaini rezultati konacnog algoritma	/0

7 Zaključak	81
Bibliografija	83
Popis slika	87
Popis tablica	89
Abstract	91
Životopis	93

# Zahvale

Htio bih se zahvaliti Prof. dr. sc. Ivanu Petroviću na sveukupnoj potpori i savjetima tokom izrade diplomskog rada i studiranja. Zahvaljujem se i svojoj obitelji na beskrajnoj životnoj mudrosti i potpori. Također se zahvaljujem svim djelatnicima Zavoda za automatiku i računalno inženjerstvo na nesebičnom dijeljenu znanja i iskustva.

### Poglavlje 1

## Uvod

Ovaj diplomski rad bavi se područjem lokalizacije govornika na temelju zvučnog signala. Senzor zvuka biti će polje mikrofona određenog broja i geometrije koje će predstavljati auditorni sustav mobilnog robota. Određivanje položaja govornika u koordinatnom sustavu mobilnog robota temelji se na mjerenju razlika u vremenima dolaska signala (engl. *time difference of arrival* tj. TDOA) između mikrofona. Kako će se pokazati, točno određivanje TDOA temelj je svakog robusnog sustava za lokalizaciju i praćenje govornika.

Cilj diplomskog rada je razviti auditorni sustav za mobilnog robota, koji će biti u stanju locirati i pratiti govornika kroz vrijeme, te ga testirati na modelu Pioneer 3DX u prostorijama Zavoda za automatiku i računalno inženjerstvo.

#### 1.1 Motivacija i cilj

U svakodnevnom životu ljudi se uvelike oslanjaju na čulo sluha kao dodatno oruđe u razumijevanju svijeta oko nas. Naše akustičko okruženje organiziramo stvarajući različite audio tokove, gdje svaki tok predstavlja određenu jedinku koja proizvodi zvuk. Svaki tok sadrži percepcijske atribute koji su snažno povezani s fizičkim svojstvima izvora zvuka; kao što je fizička veličina, svojstva prostorije u kojoj se nalazimo te relativni položaj izvora zvuka od slušatelja. Svi ovi atributi audio tokova pomažu nam da se lakše snalazimo u prostoru, npr. lako možemo odrediti da li se nalazimo u pećini ili sobi te gdje se nalazi govornik koji započinje razgovor s nama. Sluh kao jedno od tradicionalnih pet osjetila odlično nadopunjava ostala osjetila, kao vid na primjer, tako što je svesmjeran, na njega ne utječe osvijetljenost kao niti fizičke prepreke. Životinjama koje nisu grabežljivci mogućnost određivanja udaljenosti i azimuta (akustička lokacija) prijetnje sluhom, od životne je važnosti te često prethodi vizualnoj potvrdi. Iako je kroz tisuće godina razvoja čovjek izgubio potrebu za stalnim osluškivanjem okoline radi bijega od grabežljivca, danas je još uvijek sposoban s točnošću od nekoliko stupnjeva odrediti azimut i elevaciju izvora zvuka, dok u određivanju udaljenosti životinjski svijet znatno prednjači.

Zbog ovih jedinstvenih svojstava sluha kao senzora, od velike i praktične je važnosti bolje razumjeti kako ono funkcionira te pronaći sve bolja i bolja inženjerska rješenja za uvođenje osjetila sluha (auditornog sustava) u automatizirane sustave. Problem "obdarivanja" automatiziranih sustava čulom sluha je vrlo široke prirode te naravno ovisi o području i svrsi primjene. Automatizirani sustav kojim će se baviti ovaj diplomski rad je mobilni robot. Karakteristično za ovakav sustav je što okolina nije kontrolirana (robot se može nalaziti na otvorenom, u muzeju, dnevnoj sobi) i što su resursi ograničeni (auditorni sustav mora biti malih dimenzija i malih zahtjeva za energijom). Auditorni sustav znatno bi unaprijedio mogućnost mobilnih robota da lociraju i prate osobe u svojoj okolini. Imajući dotičnu mogućnost i drugi sustavi na mobilnom robotu, poput prepoznavanja govora ili prepoznavanja govornika, znatno bi bili poboljšani jer bi omogućavali prostornu diskriminaciju izvora zvukova.

Cilj ovog diplomskog rada je rješavanje konkretno definiranog problema s konkretnim odgovorom; problem akustičke lokalizacije izvora zvuka. U većini slučajeva, fizički položaj izvora zvuka je prirodan odgovor na pitanje lokacije. Međutim, što ako se izvor nalazi u sobi na kraju hodnika? Da li onda lokaciju izvora definirati kao hodnik ili drugu sobu? Koliko je iz aspekta mobilnog robota jedan odgovor lošiji od drugog? Ovaj diplomski rad neće se baviti dotičnim pitanjima, već će se ograničiti na slučaj kada postoji izravan put bez prepreka od jednog izvora zvuka do mobilnog robota. Kako će se pokazati i ovako definirani problem više je nego izazovan i zanimljiv.

#### 1.2 Pregled rada

U poglavlju 2 dane su osnove teorije lokalizacije; predstavljena je ideja prostorno-vremenskih signala, na koji način se šire kroz prostor i kakav utjecaj to ima na snimanje mikrofonima. Ispostavlja se da snimanje prostornih svojstava zvučnog signala daje mogućnost određivanja lokacije zvuka. Također su razmotrena dva bitna čimbenika prilikom snimanja, a to su odjek i šum prostorije.

U poglavlju 3 opisane su najčešće metode estimacije razlike u vremenu dolaska između signala (engl. *Time Difference of Arrival, TDOA*). Odabrana je metoda koja se temelji na računanju kroskorelacije te su u nastavku predstavljena poboljšanja dotične metode različitim funkcijama. Na kraju poglavlja dan je pregled algoritma.

U poglavlju 4 prikazano je kako se iz TDOA mjerenja određuje azimut izvora zvuka. Na kraju poglavlja dani su rezultati eksperimenta određivanja azimuta s dva mikrofona. Napravljen je ukupno 21 test za 7 različitih kuteva s tri različite pobude.

U poglavlju 5 dane su osnove hiperboličke multilateracije, tj., određivanja položaja govornika metodom sjecišta hiperbola. Pokazano je na koji način TDOA mjerenja para mikrofona definiraju hiperbolu te kako pronaći eksplicitno jednoznačno rješenje ovih nelinearnih jednadžbi Bucherovom metodom. Na kraju poglavlja su dani rezultati Monte Carlo simulacija Bucherove metode te je objašnjeno zašto ona nije prikladna za ovaj rad.

U poglavlju 6 dan je konačan algoritam lokalizacije i praćenja govornika. Objašnjena je predložena metoda za određivanje lokacije govornika metodom sjecišta asimptota hiperbola te je ukratko dana usporedba dvaju različitih geometrija polja mikrofona; kvadratne i Y geometrije. Također je i objašnjen odabrani algoritam za praćenje govornika temeljen na metodi rekurzivnih najmanjih kvadrata (RLS). Na kraju poglavlja dani su rezultati određivanja azimuta govornika poljem mikrofona na mobilnom robotu.

### Poglavlje 2

### Teorija lokalizacije

U jednokanalnim tehnikama audio sučelja radi se s valnim oblicima koji su kontinuirana funkcija vremena. Cilj višekanalnog snimanja zvuka je iskorištavanje i strukture propagirajućih valova, tj., prostornih i vremenskih svojstava zvuka prilikom digitalne obrade signala. Snimljeni signali su stoga funkcije prostora i vremena te se stoga i nazivaju prostornovremenskim signalima.

Ovo poglavlje je organizirano kako slijedi: u poglavlju 2.1 predstavlja se ideja prostorno-vremenskog signala i na koji način ga je najbolje matematički modelirati. U poglavlju 2.2 daje se matematička formulacija prostorno-vremenskog signala i raspravlja se razlika između planarnog i sfernog vala te na koji način oba utječu na lokalizaciju govornika. U poglavlju 2.3 predstavljaju se šum i odjek prostorije kao druga dva čimbenika koja znatno otežavaju lokalizaciju.

#### 2.1 Prostorno-vremenski signali

Prostorno-vremenski signali su determinističke funkcije položaja i vremena. Njihova svojstva su uvjetovana fizikom, naravno, pogotovo valnom jednadžbom. Isto kao što se odziv nekog sustava može odrediti pomoću vremenskog impulsnog odziva isto tako se i propagacija valova u nekom akustičkom prostoru može modelirati pomoću prostornovremenskog impulsnog odziva. Vrlo često deterministički model prostornovremenskog signala ne može adekvatno opisati akustičke signale, jer vrlo teško složeni akustički signali mogu biti opisani pomoću jedinstvene funkcije koja svakom vremenskom trenutku pridjeljuje jedinstvenu numeričku vrijednost. Osim toga, pronaći analitička rješenja prostornovremenskih impulsnih odziva prostorija je više nego zahtjevno. U takvim situacijama puno je prikladnije promatrati signale sa stohastičkog stajališta, tj. modelirati ih kao slučajne procese. Usprkos neadekvatnosti deterministčkog pristupa, na njihovom primjeru lakše će se približiti ideje i problematika u obradi signala pomoću polja mikrofona.

#### 2.2 Zvučno polje kao prostorno-vremenski signal

Prostorno vremenski signal označava se kao  $s(\mathbf{p}, t)$  gdje  $\mathbf{p}$  predstavlja položaj s kojeg se promatra signal, a t označava vrijeme. Vektor položaja  $\mathbf{p}$  je vektor u 3-D kartezijevom koordinatnom sustavu s koordinatama (x, y, z). Sferne koordinate su označene s  $(r, \theta, \phi)$ , gdje je r polumjer,  $\theta$ kut elevacije, a  $\phi$  azimut. Slika 2.1 prikazuje vezu između kartezijevih i sfernih koordinata.



Slika 2.1: Veza između kartezijevih i sfernih koordinata

Ovakav signal predstavlja zapravo polje, čija svojstva se mijenjaju s vremenom i s obzirom na položaj s kojeg ga promatramo. Tako možemo stajati na istom mjestu i kroz vrijeme promatrati signal, možemo fiksirati vrijeme i promatrati kako se signal mijenja s promjenom položaja promatranja ili se može kroz vrijeme mijenjati položaj i promatrati signal. Npr. u slučaju polja mikrofona upravo se radi posljednje, jer se signal uzorkuje kako vrijeme prolazi, ali također uzorkujemo i njegova prostorna svojstva s međusobno razmještenim mikrofonima. U našem slučaju prostorno-vremenski signal  $s(\mathbf{p}, t)$  opisuje zvučni tlak akustičkog polja kao funkciju položaja promatranja  $\mathbf{p}$  i vremena t. Pretpostavke na medij kroz koji se širi zvučni val su sljedeće:

- Homogenost osigurava konstantu brzinu širenja kroz prostor i vrijeme,
- Nema disperzija disperzije se događaju u nelinearnim medijima gdje širenje ovisi o frekvenciji i amplitudi zvučnog vala,
- Nema gubitaka odnosno medij ne utječe na smanjenje amplitude tokom gibanja vala kroz prostor.

Uz navede pretpostavke širenje zvučnog vala u prostoru određeno je skalarnom linearnom valnom jednadžbom [15]:

$$\nabla^2 s(\mathbf{p}, t) = \frac{1}{c^2} \frac{\partial^2 s(\mathbf{p}, t)}{\partial t^2},$$
(2.1)

gdje je  $\nabla^2$  Laplacijan drugog reda, a *c* brzina širenja vala. Jednostavnije rečeno, jednadžba (2.1) opisuje širenje zvuka u vakuumu, a u našem slučaju i širenje od izvora zvuka do senzora, tj. mikrofona. Jedno rješenje jednadžbe (2.1) je monokromatski ravni val koji se može općenito zapisati kao:

$$s(\mathbf{p},t) = A \cdot e^{j(\omega_0 t - \mathbf{k}_0^{\mathsf{l}} \mathbf{p})},\tag{2.2}$$

gdje je *A* amplituda, a  $\omega_0$  frekvencija monokromatskog ravnog vala. Vektor **k**<sub>0</sub> predstavlja valni vektor koji pokazuje u smjeru širenja vala, a amplituda vektora **k**<sub>0</sub> se još naziva i valni broj te je definirana kao:

$$|\mathbf{k}_0| = \frac{\omega_0}{c} = \frac{2\pi}{\lambda_0},\tag{2.3}$$

gdje je  $\lambda_0$  valna duljina. Valni vektor je u biti prostorna frekvencija signala i mjeri se u [rad/m] s obzirom na frekvenciju  $\omega_0$  te zapravo prikazuje koliko valnih duljina emitiranog signala stane u jedan metar. U slučaju ravnog vala, valna fronta je ploha konstantne faze, koja je općenito okomita na smjer širenja vala. Za ravni val vrijedi da ima konstantnu fazu u svim ravninama okomitim na smjer propagiranja vala u danom vremenu t. Skalarni produkt  $\mathbf{k}_{0}^{T}\mathbf{p}$  predstavlja kašnjenje signala s obzirom na ishodište koordinatnog sustava.

Drugo rješenje jednadžbe (2.1) jest monokromatski sferni val kod kojeg je faza konstanta na sfernim plohama. U ovom slučaju potrebno je prikazati signal u sfernom koordinatnom sustavu (r,  $\theta$ ,  $\phi$ ) te rješenje ima sljedeći oblik:

$$s(r,t) = \frac{A}{r} \cdot e^{j(\omega_0 t - |\mathbf{k}_0|r)}.$$
(2.4)

Za razliku od ravnog vala, amplituda monokromatskog sfernog vala smanjivati će se obrnuto proporcionalno s *r*.

Jednadžbe (2.2) i (2.4) potrebne su da bi se modelirala dva različita uvjeta prilikom uzorkovanja signala. Nadolazeći val se modelira kao sferni ako je položaj promatranja blizu izvora, tada se nalazimo u tzv. *near-field* uvjetima. Za velike udaljenosti mjesta promatranja od izvora, zakrivljenost vala se smanjuje te se val može aproksimirati ravnim valom, tada se nalazimo u tzv. *far-field* uvjetima. Slika 2.2 prikazuje dotične dvije situacije te kako pretpostavka ravnog vala može unositi pogrešku u modeliranju. Naime, ako pretpostavimo širenje ravnim valom u nearfield uvjetima, događa se kao da je zvučni val stigao za  $\Delta t$  ranije. Neovisno o pogrešci koju far-field pretpostavka može unijeti u model, često se upotrebljava u metodama lokalizacije jer ovakvo pojednostavljenje znatno olakšava proračun smjera dolaska zvuka.

Za širokopojasne signale, poput govora, teško je reći kada nastupa dotični prijelaz uvjeta i što je točno velika udaljenost od mikrofona. Općenito se uzima pretpostavka da far-field uvjeti nastupaju kada je udaljenost od mikrofona dovoljno veća od razmaka između mikrofona. Međutim, za monokromatske signale prijelaz se može analitički definirati:

$$r > \frac{2d^2}{\lambda_0},\tag{2.5}$$

gdje je *r* udaljenost do mikrofona, *d* udaljenost između senzora te  $\lambda_0$  valna duljina monokromatskog vala [14].

Kao što je ranije konstatirano u slučaju govornika teško je odrediti gdje dolazi do prijelaza jer je govor daleko od monokromatskog vala. Čak i za periodičke komponente govora poput samoglasnika kada se može pričati o valnoj duljini, ona opet uvelike ovisi o govorniku. Frekvencijski raspon osnovnog harmonika glasa odraslog muškarca iznosi od 85-155 Hz, a



Slika 2.2: Širenje ravnog i sfernog vala iz točkastog izvora

žene od 165-255 Hz [1]. Praksa je u telefonskim sustavima i sustavima za obradu govora da se glas filtrira pojasno-propusnim filtrom graničnih frekvencija  $f_d$  = 300 Hz i  $f_g$  = 3000 Hz. Iako je donja granica iznad osnovne frekvencije većine govornika, viši harmonici sadrže dovoljno informacija o spektralnom sastavu govornog signala, tako da se još uvijek može prepoznati osoba s kojom se priča na telefonu, kao što i sustavi za prepoznavanje govora mogu s određenom točnošću prepoznavati govor. Ako se ograničimo na navedeni telefonijski raspon frekvencija, za razmak mikrofona d = 50 cm far-field uvjeti za najbolji slučaju (f = 300 Hz) nastupaju pri  $r \approx 40$  cm, a za najgori slučaj f = 3000 Hz pri  $r \approx 4.4$  m. Pošto se govornik očekuje u okolini mobilnog robota barem jedan korak dalje od samog robota, može se zaključiti da će zapravo interakcija čovjekrobot biti mješavina near-field i far-field uvjeta. Za spomenutu udaljenost interakcije, far-field pretpostavka je često dostatna te se u praksi najčešće pridržava "pravila palca" i udaljenost između mikrofona se drži između 30 i 60 cm.

#### 2.3 Šum i odjek

U sustavima lokalizacije govornika glavni uzroci netočnih mjerenju su šum i odjek koji se javljaju u prostoriji. Osim termičkog šuma, koji mogu uzrokovati mikrofoni i ostala elektronička oprema, šum dolazi i od okoline (npr. ventilatori okolnih kompjutera ili motori gibajućeg robota). Pošto se većina metoda za lokalizaciju temelji na uspoređivanju signala na različitim mikrofonima ili njihovim zbrajanjem te traženje maksimuma energije, dok je god odnos signal-šum dovoljno velik, okolna buka neće predstavljati veliki problem u lokalizaciji.

Odjek prostorije predstavlja puno veći problem u točnoj lokalizaciji. Širenje zvučnog signala u zatvorenom prostoru je uvijek popraćeno s refleksijama. Stoga signal koji se dobije na mikrofonu ne sadrži samo direktnu komponentu od izvora do mikrofona, već i mnoštvo verzija originalnog signala reflektiranog od zidova, poda i različitih prepreka. U kolikoj mjeri se javlja odjek u nekoj prostoriji definirano je vremenom odjeka prostorije  $T_{60}$  koje mjeri koliko je vremena potrebno da se signal nakon prestanka emitiranja priguši za 60 dB. Poznata Sabineova formula daje vrijeme odjeka:

$$T_{60} = 0.161 \frac{V}{\sum_{i} a_i S_i},$$
(2.6)

gdje je V volumen prostorije, a  $a_i$  faktor prigušenja površine  $S_i$ .

Ako se pogleda slika 2.3 može se vidjeti da impulsni odziv jedne tipične prostorije se sastoji od direktnog signala, ranih i kasnih refleksija. Rane refleksije koje se javljaju netom nakon direktnog signala najveći su uzrok netočnih mjerenja, jer rano reflektirani signal je visoko koreliran s originalnim te ih je teško razlikovati. Upravo zbog velikog utjecaja odjeka na točnost estimacije, robusnost s obzirom na odjek je čest parametar prema kojem se različite metode lokalizacije uspoređuju.

Dobro je spomenuti da je odjek u nekim područjima, poput glazbe, poželjan efekt jer daje slušaocu različit akustički doživljaj. Zanimljivo područje akustike je mjerenje impulsnih odziva različitih koncertnih hala te pokušaj digitalne reprodukcije glazbe s istim efektom. Stoga, odjek ne mora nužno biti loš faktor u snimanju, već ovisi o području primjene.



Slika 2.3: Skica impulsnog odziva prostorije

#### Poglavlje 3

## **TDOA** estimacija

Estimacija razlike u vremenu dolaska signala nalazi primjenu u mnogim područjima, od radara, sonara, seizmologije, geofizike pa do bežičnih komunikacija. Ona je prvi korak koji nadalje daje informacije algoritmima za otkrivanje, lokalizaciju i praćenje izvora koji mogu biti akustičke (tj. mehaničke) ili elektromagnetske prirode. Ovo područje je znatno napredovalo posljednjih nekoliko desetljeća i nastavlja istim tempom u svrhu stvaranja sustava koji su robusni s obzirom na šum i odjek (koji predstavlja glavni problem u točnoj estimaciji TDOA).

Ovo poglavlje je organizirano kako slijedi: u poglavlju 3.1 objašnjava se što je točno TDOA i koja je ideja iza njegovog mjerenja s prostorno razmještenim senzorima. U poglavlju 3.2 objašnjene su različite metode estimacije TDOA, njihove prednosti i nedostatci te koja je metoda izabrana i zašto. U poglavlju 3.3 predstavljena je matematička formulacija odabrane metode, definicija kroskorelacije te kojim izmjenama se može poboljšati estimacija TDOA. U poglavlju 3.4 dan je kratak pregled razvijenog algoritam za estimaciju TDOA,

#### 3.1 TD estimacija

Estimacija TDOA je mjerenje vremena koje je potrebno valnoj fronti da proputuje udaljenost između dva prostorno razmještena senzora. Ovakva mjerenja najčešće se primjenjuju u pasivnim sonarima i poljima mikrofona. Estimacija TDOA pripada području estimacije vremena kašnjenja (engl. *Time Delay Estimation*, TDE) i druga je grana jedne općenite podjele kojoj pripada i estimacija vremena dolaska (engl. *Time of Arrival*, TOA). Dakle, ovisno o načinu primjene, TDE može se podijeliti na:

- TOA estimaciju mjerenje vremena kašnjenja između odaslanog signala i primitka njegovog odjeka; najčešće se primjenjuje u aktivnim sustavim poput radara i aktivnog sonara,
- TDOA estimaciju mjerenje vremena kašnjenja između referentnog i nekog drugog senzora.

Iako postoji intrinzična veza između ova dva tipa estimacije, njihova problematika ih naveliko razlikuje. Na primjer, kod slučaja radara, mjerenje TOA je olakšano činjenicom da je poznat referentni signal te je povratni signal samo zašumljena verzija referentnog. Ovakvim pristupom TOA se može estimirati jednim senzorom filtrirajući dolazni signal i uspoređujući dotični s referentnim. Estimacija TDOA je ponešto složenije prirode jer nije dostupan eksplicitni referentni signal, pojavljivanje i statistika izvora su nepoznati, te se najčešće razlika u vremenu dolaska određuje uspoređujući signale od dva (ili više) prostorno razdvojena senzora. Ovaj postupak se još često naziva i prostornim uzorkovanjem. U ovome radu senzore će činiti polje mikrofona, koje je očigledno pasivan sustav, te će se baviti problematikom određivanja TDOA.

#### 3.2 Metode određivanja TDOA

Estimacija TDOA bio bi poprilično lagan zadatak kada bi signali na mikrofonima bili jednostavno zakašnjele i skalirane verzije originalnog signala. U praksi signal je često zašumljen ambijentalnom bukom, jer se ipak robot kreće u svakodnevnim prostorima. Osim šuma svaki promatrani signal u biti sadrži i više prigušenih i zakašnjelih replika originalnog signala zbog refleksija od prepreka i zidova te unosi drugi problem u estimaciji TDOA, odjek. Zbog odjeka, signal koji dolazi do mikrofona, ne sadrži samo direktnu komponentu već jeku i spektralne distorzije koji dodatno iskrivljuju sliku izvornog signala. Zbog šarolike problematike postoji i široka lepeza algoritama za određivanje TDOA koji se tradicionalno dijele u tri grupe: tzv. *beamforming* metode, metode koje se temelje na spektralnoj estimaciji signala i na metode koje se temelje na estimaciji TDOA. Detaljniji pregled i usporedba metoda može se naći u [2] i [6]. Prva grupa metoda, beamforming metode, odnose se na pasivna polja mikrofona gdje se lokacija govornika određuje iz zbrojenih, filtriranih i pomaknutih signala s različitih mikrofonskih kanala, otkuda i naziv beamforming, tj. snopiranje. Temeljna ideja ovih metoda jest da se signali iz svih mikrofona zakasne i zbroje za sva moguća prije izračunata kašnjenja signala te se lokacija određuje kao maksimum energije izlaza ovakvog sustava. Drugačije rečeno, polje mikrofona se usmjerava na sve moguće lokacije te se govornik očekuje tamo gdje je energija signala najveća. Prostor oko polja mikrofona se najčešće podijeli u ćelije, po željenoj rezoluciji, te se za svaku izračunaju pripadajuća kašnjenja. Na ovaj način se uobičajeno radi baza kašnjenja.

Ovu grupu metoda karakterizira mogućnost lokalizacije više govornika, robusnost, ali i računska zahtjevnost.

Drugoj grupi metoda pripadaju sve metode koje za proračun lokacije koriste prostorno-spektralnu korelacijsku matricu signala. Najpoznatiji algoritmi iz ove skupine su: autoregresivno (AR) modeliranje, spektralna estimacija s minimalnom varijancom i različite metoda koje se temelje na analizi svojstvenih vrijednosti (npr. MUSIC algoritam).

Ovu grupu metoda karakterizira također lokalizacija više govornika, manja računska zahtjevnost od beamforming metoda, ali su većinom razvijene za uskopojasne signale i osjetljive su na odjek. Navedeni nedostatci se mogu nadići određenim postupcima, ali na uštrb računske prednosti naspram beamforming metoda.

Posljednjoj grupi metoda pripadaju oni algoritmi koji određuju lokaciju govornika iz skupa estimata kašnjenja između različitih parova mikrofona. Osnovna ideja je podijeljena u dva koraka: prvo se računa kašnjenje između različitih prostorno razmještenih mikrofona (TDOA) te se potom iz kašnjenja za svaki par mikrofona dobivaju hiperbole čije sjecište određuje lokaciju govornika. Ove metode se razlikuju po načinu na koji određuju TDOA (najčešće se primjenjuje kroskorelacijski postupak) te kojim načinom iz TDOA određuju lokaciju govornika. Ovaj postupak je detaljnije pojašnjen u poglavlju 3.4.

Zbog svoje elegancije i male računske složenosti ova grupa algoritama je poprimila najveću pažnju u pasivnoj lokalizaciji. Manje je robusna od beamforming metoda i ne omogućava lokalizaciju više govornika, ali je zato i puno manje računski zahtjevna.

Pošto mobilni robot predstavlja platformu s ograničenom računskom snagom, od tri grupe algoritama odabrana je treća koja se temelji na TDOA estimaciji jer je računski najmanje zahtjevna i kako će se vidjeti uz određene pretpostavke daje optimalni estimator.

#### 3.3 Generalizirana kroskorelacija (GCC)

Da bi se odredio TDOA između dva signala, potrebno je definirati funkciju koja će dati globalni maksimum na mjestu točnog vremena kašnjenja. Kroskorelacija je najlogičniji izbor, pošto imamo dva prostorno razmještena mikrofona koja se u idealnom slučaju razlikuju jedino u fazi. Signal s(t) emitiran iz udaljenog izvora i koji se promatra na dva prostorno različita mjesta, može se modelirati na sljedeći način:

$$s_{1}(t) = s(t + t_{1}) + n_{1}(t),$$
  

$$s_{2}(t) = s(t + t_{2}) + n_{2}(t),$$
(3.1)

gdje su s(t),  $n_1(t)$  i  $n_2(t)$  Gaussovi, realni, stacionarni u širem smislu slučajni procesi te je zanemareno smanjenje amplitude s udaljenošću, a  $t_1$  i  $t_2$  predstavlja vrijeme koje je potrebno da signal s(t) stigne do prvog i drugog mjesta promatranja. Također se pretpostavlja da je signal s(t) nekoreliran sa šumom  $n_1(t)$  i  $n_2(t)$ . Iako je glas daleko od stacionarnog procesa, ta pretpostavka vrijedi za kratke konačne intervale promatranja. Ako se  $\tau_{12}$ definira kao  $\tau_{12} = t_1 - t_2$  onda ono predstavlja TDOA između signala  $s_1$ i  $s_2$ , tj., ono vrijeme koje je potrebno da signal stigne od prvog mjesta promatranja do drugog. U našem slučaju na dotičnim mjestima promatranja biti će mikrofoni koji će uzorkovati signal s(t), tako će signal prvog mikrofona u diskretnoj domeni biti  $x_{m_1}[n]$  (referentni mikrofon), a signal drugog mikrofona  $x_{m_2}[n - \tau_{12}]$ . Kroskorelacijska funkcija za slučajne procese definirana je na sljedeći način:

$$R_{x_1x_2}(\Delta \tau) = E\{x_1(t)x_2(t-\tau_{12})\},\tag{3.2}$$

gdje E{.} predstavlja operator očekivanja. Zbog konačnog vremena promatranja kroskorelacija  $R_{x_1x_2}$  mora se estimirati. Promatramo konačan blok podataka duljine L te kroskorelaciju definiramo sljedećim izrazom:

$$\hat{R}_{ij}(\tau) = \sum_{n=0}^{L-1} x_{m_i}[n] x_{m_j}[n-\tau_{ij}], \qquad (3.3)$$

gdje je  $x_{m_i}$  signal s *i*-tog mikrofona, a  $\tau_{ij}$  kašnjenje signala u uzorcima. Kako je rečeno ranije, kroskorelacija dana izrazom (3.3) imati će globalni maksimum upravo u kašnjenju između dva primljena signala. Računajući kroskorelaciju preko (3.3) složenost bi bila  $O(N^2)$ . Jedno od najprimamljivijih svojstava kroskorelacije je mogućnost računanja u frekvencijskoj domeni što smanjuje složenost na  $O(N \log_2 N)$ . Izraz (3.3) u frekvencijskoj domeni ima sljedeći oblik:

$$\hat{R}_{ij}(\tau) = \sum_{k=0}^{L-1} X_{m_i}(k) X_{m_j}^*(k) e^{j2\pi \frac{k\tau}{L}},$$
(3.4)

gdje je  $X_{m_i}$  diskretna Fourierova transformacija (DFT) od  $x_{m_i}[n]$  te (.)\* predstavlja operator kompleksnog konjugiranja. Dobro je primijetiti sličnost kroskorelacije s konvolucijom. Ono što konvolucija u vremenskoj domeni radi jest zrcaljenje jednog signala i prelaženje preko drugog, množeći i sumirajući uzorke. Za razliku od konvolucije, kroskorelacija ne zrcali signal prije prelaženja preko drugog signala. Konvoluciji signala u frekvencijskoj domeni odgovara množenje signala, a kroskorelaciji množenje s jednim kompleksno-konjugiranim (zato nema zrcaljenja) signalom. Dakle, postoji samo manja matematička razlika između ove dvije operacije, no razlika u rezultatu i primjeni je velika.

Veliki nedostatak kroskorelacije dane izrazom (3.4) je taj što susjedni uzorci su statistički visoko korelirani, što rezultira širokim i zaobljenim maksimumima. Da bi se smanjio utjecaj ovoga problema, pribjegava se otežavanjem funkcije u frekvencijskoj domeni.

#### 3.3.1 Otežavanje funkcije kroskorelacije

Jedan od načina na koji se može riješiti problem širokih maksimuma, jest da se prije inverzne diskretne Fourierove transformacije (IDFT) signal prikladno filtrira. Pokazano je u [17] da za kroskorelaciju signala  $x_{m_1}$  i  $x_{m_2}$ vrijedi sljedeća relacija:

$$R_{x_{m_1}x_{m_2}}(\tau) = R_{s_1s_1}(\tau) * \delta(\tau - \tau_{12}), \tag{3.5}$$

odnosno, delta funkcija koja predstavlja točno kašnjenje je razmazana autokorelacijom samoga signala. U slučaju da je signal s(t) bijeli šum, onda će u izrazu (3.5) ostati samo delta funkcija na točnom kašnjenju. Dakle, treba filtrirati signale tako da se izbijeli krosspektar signala  $x_{m_1}$  i  $x_{m_2}$  prije računanja IDFT-a. Najčešći izbor je PHAT težinska funkcija s kojom kroskorelacija poprima sljedeći oblik:

$$\hat{R}_{ij}^{\text{PHAT}}(\tau) = \sum_{k=0}^{L-1} \frac{X_{m_i}(k) X_{m_j}^*(k)}{|X_{m_i}(k)| |X_{m_j}^*(k)|} e^{j2\pi \frac{k\tau}{L}},$$
(3.6)

$$\psi_{\text{PHAT}} = \frac{1}{|X_{m_i}(k)||X_{m_i}^*(k)|}.$$
(3.7)

Ono što se dobiva s (3.7) jest izbjeljivanje spektra te oštriji maksimum kroskorelacije kod točnog vremena kašnjenja. Slika 3.1 pokazuje rezultat otežavanja u idealnom slučaju (filtrirana kroskorelacija ima delta funkciju u točnom vremenu kašnjenja). Idealan slučaj na slici 3.1 je u biti autokorelacija snimljenog signala, što se i vidi jer kroskorelacija daje kašnjenje u nuli, tj. nema fazne razlike među signalima.

Kroskorelacija dana izrazom (3.6) obično se naziva generaliziranom kroskorelacijom (GCC), jer postoji puno različitih otežavajućih funkcija koje se mogu upotrijebiti umjesto PHAT funkcije. Pregled nekih se može naći u [17]. Također je pokazano u [17] da izraz (3.6), ako signal imam konstantan odnos signal-šum kroz cijeli spektar, predstavlja Maximum Likelihood estimator (MLE) vremena kašnjenja. Detaljan izvod se može naći i u [23].

#### **3.3.2 PHAT-***β* težinska funkcija

Glavni nedostatak PHAT funkcije je što ona otežava sve frekvencije bez obzira na odnos signal-šum. Prilikom eksperimenata, nisu dobiveni dobri rezultati računajući kroskorelaciju pomoću (3.6), već je trebalo smanjiti utjecaj PHAT funkcije na određeni način. To nas je odvelo ka tzv. PHAT- $\beta$  težinskoj funkciji predloženoj u [10]. PHAT- $\beta$  težinska funkcija ima sljedeći oblik:

$$\psi_{\text{PHAT}-\beta} = \frac{1}{(|X_{m_i}(k)||X_{m_i}^*(k)|)^{\beta}},$$
(3.8)

gdje je 0 <  $\beta$  < 1 podešavajući parametar. Razlog zašto PHAT funkcija pokazuje loše rezultate pokazano je i objašnjeno u [10]. Naime, glas promatran na kratkim vremenskim intervalima može imati karakteristike kratkopojasnog i širokopojasnog signala. Npr. u riječi "šum", suglasnik "š" kada se promotri frekvencijski ima vrlo širok raspon i djeluje poput šuma, dok samoglasnik "u" ima uski frekvencijski raspon. U [10] predložen je iznos parametra  $\beta$  = 0.5 kao dobar kompromis za široko i uskopojasne signale te je taj iznos i korišten u eksperimentima te je



Slika 3.1: Kroskorelacija sa i bez PHAT težinske funkcije

također pokazano da dotična modifikacija ne utječe na algoritam u smislu degradacije zbog šuma ili odjeka u usporedbi s PHAT funkcijom. Naravno, ima mjesta i za daljnje poboljšanje ovoga algoritma. Moguće je uvesti masku koja se temelji na odnosu signal-šum kako je to predloženo u [12]. Također je u dotičnu masku moguće uvesti modeliranje odjeka te na taj način učiniti algoritam još robusnijim [20].

#### 3.4 Pregled algoritma za određivanje TDOA

Slika 3.2 prikazuje blok shemu algoritma za određivanje TDOA. Izvor signala s(t) u nekom trenutku emitira zvuk. Izvor zvuka na slici je modeliran kao točkasti izvor sfernog vala. Ako su mikrofoni dovoljno daleko udaljeni od izvora zvuka, val se može na njihovom mjestu smatrati ravnim valom jer se s udaljenošću zakrivljenost smanjuje (far-field pretpostavka). Mikrofoni uzorkuju zvučni val i izlaz svakog mikrofona je predstavljen kao diskretni signal  $x_{m_1}[n]$  i  $x_{m_2}[n]$ . Nad svakim signalom se vrši DFT,  $O(N \log_2 N)$ , jedan od njih se kompleksno konjugira te se signali pomnože,  $O(N\log_2 N+N)$ . Prije IDFT-a radi se otežavanje s PHAT funkcijom (realno dijeljenje) te konačna kompleksnost algoritma iznosi  $O(3/2N\log_2 N+N)$ 



Slika 3.2: Blok dijagram algoritma estimacije TDOA

Pošto je određen algoritam estimacije TDOA, prelazi se na drugi korak, a to je estimacija smjera dolaska signala.

#### Poglavlje 4

### DOA estimacija

Krajnji cilj svakog algoritma lokalizacije je, naravno, određivanje smjera dolaska signala (engl. *direction of arrival*, DOA). U seizmologiji određivanje smjera dolaska može pomoći pri traženju izvora potresa, u praćenju morskih životinja često se pomoću hidrofona određuju položaji kitova u moru itd. Prije nastanka radara, DOA metode su se koristile i pri lokalizaciji aviona pomoću mikrofona, ali ipak najveći utjecaj su imale u traženju podmornica pomoću sonara. Danas, metode hiperboličke mulitlateracije koje se temelje na TDOA, smatraju se metodama budućnosti lokalizacije te je njihova upotreba često popraćena i nazivom vrijeme poslije radara.

Ovo poglavlje je organizirano kako slijedi: u poglavlju 4.1 objašnjeno je kako se iz TDOA mjerenja računa DOA, tj. azimut, zvučnog signala. U poglavlju 4.2 objašnjen je detektor govorne aktivnosti koji se koristi u blokovskoj obradi snimljenog signala. U poglavlju 4.3 dani su rezultati eksperimenta koji s dva mikrofona koji su urađeni na Zavodu. U poglavlju 4.11 analizirani su dobiveni rezultati eksperimenta.

#### 4.1 Estimacija smjera dolaska

Kako je objašnjeno u poglavlju 3, estimacija TDOA između mikrofona *i* i mikrofona *j*, računa se nalazeći maksimum kroskorelacije (3.6):

$$\tau_{ij} = \arg\max_{\tau} \hat{R}_{ij}^{\text{PHAT}}(\Delta \tau).$$
(4.1)

Nakon što je izračunata estimacija TDOA, nizom geometrijskih postupaka

može se izračunati DOA, tj. azimut, izvora zvuka. Pretpostavka je da su dimenzije polja mikrofona puno manje od udaljenosti izvora, što daje far-field uvjete, te se pretpostavljaju ravni valovi. Slika 4.2 prikazuja slučaj dva mikrofona koja se nalaze u far-field okruženju. Preko zakona kosinusa lagano se dobije izraz za računanje azimuta:

$$\varphi = \pm \arccos\left(\frac{c\Delta\tau_{ij}}{d}\right),\tag{4.2}$$

gdje je *c* brzina zvuka, *d* razmak između mikrofona, a  $\varphi$  azimut. Sada je važno uočiti sljedeću činjenicu, koja se nameće u izrazu (4.2) preko ± znaka; pomoću dva mikrofon moguće je odrediti DOA, ali s nesigurnošću da li se izvor nalazi ispred ili iza para mikrofona. Ovaj problem se rješava uvođenjem više od 2 mikrofona, ali o tome će biti više rečeno u poglavlju 6.



Slika 4.1: Računanje DOA iz TDOA

Imajući algoritam za određivanje TDOA te način računanja DOA, sljedeći logičan korak je bila eksperimentalna provjera pomoću para mikrofona. No prije je trebalo realizirati detektor govorne aktivnosti.
#### 4.2 Detektor govorne aktivnosti

Da bi algoritam mogao razlikovati govornu aktivnost od pozadinske buke, realiziran je elegantan detektor govorne, odnosno zvučne aktivnosti. Ovaj modul se primjenjuje samo kod blokovske obrade. Sofisticiraniji detektori govorne aktivnosti su složeno područje i može biti rad sam za sebe, te je u ovom radu realiziran detektor na principu mjerenja energije signala koji se temelji na Parsevalovoj relaciji:

$$E_s = \frac{1}{N} \sum_{k=0}^{N-1} |X[k]|^2$$
(4.3)

Ako je energija signala  $E_s$  veća od prije definiranog praga T, onda se smatra da je detektirana govorna aktivnost te dotični blok signala ulazi u obradu.

#### 4.3 Rezultati eksperimenta DOA estimacije

Eksperimenti su napravljeni pri frekvenciji uzorkovanja  $F_s = 48$  kHz, rezoluciji 16 bita, razmak između mikrofona d = 0.5 m te je u real-time obradi korišten blok duljine L = 3072 uzorka, odnosno trajanja 64 ms. Cilj testiranja je bio proći od 0° – 90° te promotriti ponašanje i točnost sustava. Testiranje je provedeno na azimutima  $\varphi = 0^\circ, -15^\circ, -30^\circ, -45^\circ, -60^\circ, 70^\circ, -80^\circ,$ odnosno na njima približnim vrijednostima kako je govornik stajao. U ovom slučaju 0° je kut kada ravni val dolazi okomito na normalu od crte koja spaja mikrofone, a 90° je kut paralelan s dotičnom normalom mikrofona. Tako gledajući od središta mikrofona, kut se od lijeva prema desno kreće od  $-90^\circ$  do  $90^\circ$ .

Za svaki kut urađena su tri različita eksperimenta: u prvom se snimalo pucketanje prstima (slika 4.3), u drugom se izgovarala riječ "Test"(slika 4.4), a u trećem slučaju se izgovaralo "Test jedan, dva, tri, test, jedan, dva, tri" (slika 4.5). Za prva dva eksperimenta određen je DOA te apsolutne greške iz cijelog bloka podataka, dok je za treći eksperiment dodatno testirano i praćenje s blokovima podataka veličine L = 3072 bita. Da bi se bolje vidjelo ponašanje praćenja, prikazani su i histogrami estimacije DOA. Dodatno je nad blokovima podataka rađena i medijan funkcija na skupu od tri uzorka da bi se mogao ispraviti jedan outlier, a da se odziv sustava pri promjeni ne uspori.

Rezultati eksperimenta su dani u nastavku.



Slika 4.2: Položaji govornika tokom eksperimenta



Slika 4.3: Pucketanje prstima



Slika 4.5: Izgovor "Test jedan, dva, tri, test jedan, dva, tri"

# 4.4 Test 0°

Položaj govornika: (x, y) = (0, 1.2) m

## 4.4.1 Pucketanje prstima



Slika 4.6: GCC i GCC-PHAT pri pucketanju prstima na 0°

	Točno	Estimirano	$\Delta \varphi = \varphi_{\text{točno}} - \varphi_{\text{estimirano}}$
φ[°]	0	0	0
$arphi^{ ext{PHAT}}$ [°]	0	0	0
			$\Delta \tau \cdot F_s$ = broj uzoraka
			(pogreška u uzorcima)
τ [s]	0	0	0
$ au^{ ext{PHAT}}\left[ ext{s} ight]$	0	0	0

Tablica 4.1: Rezultati eksperimenta pucketanja pr<br/>stima pri $0^\circ$ 

# 4.4.2 Izgovor "Test"



Slika 4.7: GCC i GCC-PHAT pri izgovoru "Test" na 0°

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
φ[°]	0	-9.1	9.1
$arphi^{ ext{PHAT}}$ [°]	0	0	0
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au \left[ { m s}  ight]$	0 0	-2.2917e-004 0	11 0

Tablica 4.2: Rezultati eksperimenta izgovora "Test" pri $0^\circ$ 



# 4.4.3 Izgovor "Test jedan, dva, tri"

Slika 4.8: GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na 0°

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
$arphi$ [°] $arphi^{ ext{PHAT}}$ [°]	0	-0.82	0.82
	0	-0.82	0.82
,			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au ^{ m PHAT} \left[ { m s}  ight]$	0	-2.0833e-005	1
	0	-2.0833e-005	1

Tablica 4.3: Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri 0°



Slika 4.9: Praćenje pri izgovoru "Test jedan, dva, tri" na 0°



Slika 4.10: Histogram na 0°

## 4.5 Test -15°

Položaj govornika: (x, y) = (-0.28, 1.2) m

## 4.5.1 Pucketanje prstima



Slika 4.11: GCC i GCC-PHAT pri pucketanju prstima na  $-15^{\circ}$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\text{točno}} - \varphi_{\text{estimirano}}$
φ[°]	-13.1	-13.3	0.2
$arphi^{ ext{PHAT}}$ [°]	-13.1	-13.3	0.2
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight] \  au ^{ m PHAT} \left[ { m s}  ight]$	-3.3028e-004 -3.3028e-004	-3.3333e-004 -3.3333e-004	0.1 0.1

Tablica 4.4: Rezultati eksperimenta pucketanja pr<br/>stima pri $-15^\circ$ 

# 4.5.2 Izgovor "Test"



Slika 4.12: GCC i GCC-PHAT pri izgovoru "Test" na  $-15^\circ$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
φ[°]	-13.1	-4.1	-9
$\varphi^{\mathrm{PHAT}}$ [°]	-13.1	-13.3	0.2
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au \left[ { m s}  ight]$	-3.3028e-004 -3.3028e-004	-1.0417e-004 -3.3333e-004	-10.9 0.1

Tablica 4.5: Rezultati eksperimenta izgovora "Test" pri $-15^\circ$ 



# 4.5.3 Izgovor "Test jedan, dva, tri"

Slika 4.13: GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na  $-15^{\circ}$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\text{točno}} - \varphi_{\text{estimirano}}$
$arphi$ [°] $arphi^{ ext{PHAT}}$ [°]	-13.1	-14.1	1.0
	-13.1	-13.3	0.2
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au \left[ { m s}  ight]$	-3.3028e-004	-3.5417e-004	1.1
	-3.3028e-004	-3.3333e-004	0.2

Tablica 4.6: Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri $-15^{\circ}$ 



Slika 4.14: Praćenje pri izgovoru "Test jedan, dva, tri" na  $-15^\circ$ 



Slika 4.15: Histogram na $-15^\circ$ 

# 4.6 Test -30°

Položaj govornika: (x, y) = (-0.6, 1.2) m

## 4.6.1 Pucketanje prstima



Slika 4.16: GCC i GCC-PHAT pri pucketanju prstima na  $-30^{\circ}$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
φ[°]	-26.5	-26.4	-0.1
$\varphi^{\mathrm{PHAT}}$ [°]	-26.5	-26.4	-0.1
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au \left[ { m s}  ight]$	-6.5002e-004 -6.5002e-004	-6.4583e-004 -6.4583e-004	-0.2 -0.2

Tablica 4.7: Rezultati eksperimenta pucketanja pr<br/>stima pri $-30^\circ$ 

# 4.6.2 Izgovor "Test"



Slika 4.17: GCC i GCC-PHAT pri izgovoru "Test" na  $-30^\circ$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
φ[°]	-26.5	-27.3	0.8
$arphi^{ ext{PHAT}}$ [°]	-26.5	-27.3	0.8
			$\Delta \tau \cdot F_s = broj uzoraka$
			(pogreška u uzorcima)
τ [s]	-6.5002e-004	-6.6667e-004	0.8
$\tau^{\mathrm{PHAT}}\left[\mathrm{s} ight]$	-6.5002e-004	-6.6667e-004	0.8

Tablica 4.8: Rezultati eksperimenta izgovora "Test" pri $-30^\circ$ 



# 4.6.3 Izgovor "Test jedan, dva, tri"

Slika 4.18: GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na -30°

	Točno	Estimirano	$\Delta \varphi = \varphi_{\text{točno}} - \varphi_{\text{estimirano}}$
$arphi$ [°] $arphi^{ ext{PHAT}}$ [°]	-26.5	-26.4	-0.1
	-26.5	-26.4	-0.1
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au ^{ m PHAT} \left[ { m s}  ight]$	-6.5002e-004	-6.4583e-004	-0.2
	-6.5002e-004	-6.4583e-004	-0.2

Tablica 4.9: Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri $-30^{\circ}$ 



Slika 4.19: Praćenje pri izgovoru "Test jedan, dva, tri" na  $-30^\circ$ 



Slika 4.20: Histogram na  $-30^\circ$ 

## 4.7 Test -45°

Položaj govornika: (x, y) = (-0.85, 1) m

## 4.7.1 Pucketanje prstima



Slika 4.21: GCC i GCC-PHAT pri pucketanju prstima na  $-45^{\circ}$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
φ[°]	-40.4	-47	6.6
$arphi^{ ext{PHAT}}$ [°]	-40.4	-47	6.6
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
τ [s] τ <sup>PHAT</sup> [s]	-9.4135e-004 -9.4135e-004	-1.0625e-003 -1.0625e-003	5.8 5.8

Tablica 4.10: Rezultati eksperimenta pucketanja pr<br/>stima pri $-45^\circ$ 

# 4.7.2 Izgovor "Test"



Slika 4.22: GCC i GCC-PHAT pri izgovoru "Test" na  $-45^\circ$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
φ[°]	-40.4	-24.6	-15.8
$arphi^{ ext{PHAT}}$ [°]	-40.4	-44.6	4.2
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au \left[ { m s}  ight]$	-9.4135e-004 -9.4135e-004	-6.0417e-004 -1.0208e-003	-16.2 3.8

Tablica 4.11: Rezultati eksperimenta izgovora "Test" pri $-45^\circ$ 



# 4.7.3 Izgovor "Test jedan, dva, tri"

Slika 4.23: GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na  $-45^{\circ}$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\text{točno}} - \varphi_{\text{estimirano}}$
$\varphi$ [°]	-40.4	-42.4	2
$\varphi^{\text{IIAI}}[\circ]$	-40.4	-42.4	2
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au ^{ m PHAT} \left[ { m s}  ight]$	-9.4135e-004 -9.4135e-004	-9.7917e-004 -9.7917e-004	1.8 1.8

Tablica 4.12: Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri $-45^\circ$ 



Slika 4.24: Praćenje pri izgovoru "Test jedan, dva, tri" na  $-45^{\circ}$ 



Slika 4.25: Histogram na  $-45^\circ$ 

## 4.8 Test -60°

Položaj govornika: (x, y) = (-1.4, 1) m

## 4.8.1 Pucketanje prstima



Slika 4.26: GCC i GCC-PHAT pri pucketanju prstima na  $-60^{\circ}$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\text{točno}} - \varphi_{\text{estimirano}}$
φ[°]	-54.4	-61	6.6
$arphi^{ ext{PHAT}}$ [°]	-54.4	-61	6.6
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au \left[ { m s}  ight]$	-1.1828e-003 -1.1828e-003	-1.2708e-003 -1.2708e-003	4.2 4.2

Tablica 4.13: Rezultati eksperimenta pucketanja pr<br/>stima pri $-60^\circ$ 

# 4.8.2 Izgovor "Test"



Slika 4.27: GCC i GCC-PHAT pri izgovoru "Test" na  $-60^\circ$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
φ[°]	-54.4	-54.8	0.4
$arphi^{ ext{PHAT}}$ [°]	-54.4	-54.8	0.4
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au \left[ { m s}  ight]$	-1.1828e-003 -1.1828e-003	-1.1875e-003 -1.1875e-003	0.2 0.2

Tablica 4.14: Rezultati eksperimenta izgovora "Test" pri $-60^\circ$ 



# 4.8.3 Izgovor "Test jedan, dva, tri"

Slika 4.28: GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na  $-60^{\circ}$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
$\varphi [\circ]$	-54.4	-50.7	-3.7
$\varphi^{\text{PHAT}} [\circ]$	-54.4	-54.8	0.4
ΥĽΙ	0 11 1	0 110	$\Delta \tau \cdot F_s = \text{broj uzoraka}$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au ^{ m PHAT} \left[ { m s}  ight]$	-1.1828e-003	-1.1250e-003	-2.7
	-1.1828e-003	-1.1875e-003	0.2

Tablica 4.15: Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri $-60^\circ$ 



Slika 4.29: Praćenje pri izgovoru "Test jedan, dva, tri" na  $-60^{\circ}$ 



Slika 4.30: Histogram na  $-60^\circ$ 

# 4.9 Test 70°

Položaj govornika: (x, y) = (2.3, 1) m

## 4.9.1 Pucketanje prstima



Slika 4.31: GCC i GCC-PHAT pri pucketanju prstima na 70°

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
φ[°]	66.5	68.7	2.2
$arphi^{ ext{PHAT}}$ [°]	66.5	68.7	2.2
			$\Delta \tau \cdot F_s = broj uzoraka$
			(pogreška u uzorcima)
τ [s]	1.3330e-003	1.3542e-003	-1
$\tau^{\mathrm{PHAT}}\left[\mathrm{s} ight]$	1.3330e-003	1.3542e-003	-1

Tablica 4.16: Rezultati eksperimenta pucketanja pr<br/>stima pri $70^\circ$ 





Slika 4.32: GCC i GCC-PHAT pri izgovoru "Test" na 70°

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
φ[°]	66.5	41.3	25.3
$arphi^{ ext{PHAT}}$ [°]	66.5	61	5.5
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
τ [s]	1.3330e-003	9.5833e-004	18
$\tau^{\text{PHAT}}[\mathbf{s}]$	1.3330e-003	1.2708e-003	3

Tablica 4.17: Rezultati eksperimenta izgovora "Test" pri $70^\circ$ 



## 4.9.3 Izgovor "Test jedan, dva, tri"

Slika 4.33: GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na 70°

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
$\varphi [\circ]$	66.5	32	-34.5
$\varphi^{\text{PHAT}} [\circ]$	66.5	61	5 5
ΥĽΙ	00.0	01	$\Delta \tau \cdot F_s = \text{broj uzoraka}$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au ^{ m PHAT} \left[ { m s}  ight]$	1.3330e-003	7.7083e-004	27
	1.3330e-003	1.2708e-003	3

Tablica 4.18: Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri $70^\circ$ 



Slika 4.34: Praćenje pri izgovoru "Test jedan, dva, tri" na 70°



Slika 4.35: Histogram na 70°

## 4.10 Test -80°

Položaj govornika: (x, y) = (-1.8, 0.5) m

## 4.10.1 Pucketanje prstima



Slika 4.36: GCC i GCC-PHAT pri pucketanju prstima na  $-80^{\circ}$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
φ[°]	-74.5	-73.8	-0.7
$\varphi^{\text{PHAT}}[\circ]$	-74.5	-73.8	-0.7
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au \left[ { m s}  ight]$	-1.4005e-003 -1.4005e-003	-1.3958e-003 -1.3958e-003	-0.2 -0.2

Tablica 4.19: Rezultati eksperimenta pucketanja pr<br/>stima pri $-80^\circ$ 



# 4.10.2 Izgovor "Test"

Slika 4.37: GCC i GCC-PHAT pri izgovoru "Test" na  $-80^\circ$ 

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
φ[°]	-74.5	-71.1	-3.4
$arphi^{ ext{PHAT}}$ [°]	-74.5	-66.5	-8
			$\Delta \tau \cdot F_s = broj uzoraka$ (pogreška u uzorcima)
$ au \left[ { m s}  ight]  au \left[ { m s}  ight]  au ^{ m PHAT} \left[ { m s}  ight]$	-1.4005e-003 -1.4005e-003	-1.3750e-003 -1.3333e-003	-1.2 -3.2

Tablica 4.20: Rezultati eksperimenta izgovora "Test" pri $-80^\circ$ 



# 4.10.3 Izgovor "Test jedan, dva, tri"

Slika 4.38: GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na -80°

	Točno	Estimirano	$\Delta \varphi = \varphi_{\rm točno} - \varphi_{\rm estimirano}$
φ[°]	-74.5	-68.7	-5.8
$arphi^{ ext{PHAT}}$ [°]	-74.5	-66.5	-8
			$\Delta \tau \cdot F_s = broj uzoraka$
			(pogreška u uzorcima)
τ [s]	-1.4005e-003	-1.3542e-003	-2.2
$ au^{ ext{PHAT}}\left[ ext{s} ight]$	-1.4005e-003	-1.3333e-003	-3.2

Tablica 4.21: Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri $-80^\circ$ 



Slika 4.39: Praćenje pri izgovoru "Test jedan, dva, tri" na  $-80^{\circ}$ 



Slika 4.40: Histogram na  $-80^\circ$ 

#### 4.11 Analiza rezultata eksperimenata

Rezultati prikazani od poglavlja 4.4 do 4.10 dobro prikazuju problematiku koja se javlja prilikom estimacije TDOA, tj. DOA, te na koji način se ona eventualno može umanjiti.

Problem širokih maksimuma kroskorelacije dobro se vidi na slici 4.7 te kako PHAT funkcija "izoštrava" dotični maksimum. U nekim slučajevima bez obzira na široki maksimum GCC bi dao dobar rezultat, ali statistički u većini slučajevima upravo zbog toga daje veliku pogrešku (tablica 4.18). Naravno, ima i slučajeva kada se GCC pokaže boljim od GCC-PHAT funkcije, ali to su rijetki slučajevi (4.20)

Praćenje također pokazuje dobre učinke PHAT funkcije, jer smanjuje pojavljivanje outliera i varijancu distribucije DOA estimacije. Grafički se ovaj efekt dobro može vidjeti na histogramima estimacije DOA.

Krajnji zaključak eksperimenta je da estimacija DOA putem kroskorelacije i PHAT- $\beta$  funkcija daju dobre rezultate u smislu točnosti (najveća zabilježena apsolutna pogreška je 8°) i efikasnosti (elegantan i brz algoritam). Važno je još napomenuti da nepouzdanost estimacije raste kako se približava pozicijama od –90° i 90° te da je najbolja estimacija pri 0°. Razlog ovome ponašanju nalazi se u nelinearnosti arccos funkcije u izrazu (4.2). Naime, kako se sve više približava iznosima od –90° i 90° povećava se i iznos derivacije te male promjene u pogrešci estimacije TDOA izazivaju veće greške u DOA estimaciji. Na isti problem ukazano je i u [19]. Zbog diskretizacije vremenske osi pri estimaciji TDOA, postoji i, naravno, diskretizacija po stupnjevima te je moguće da se u nekim slučajevima ne može izmjeriti kut od ±75° do ±90°. O rezoluciji polja mikrofona biti će više rečeno u poglavlju 6.2.

# Poglavlje 5

# Hiperbolička multilateracija

Sljedeće pitanje koje se nameće jest, da li se može iz dobivenih mjerenja TDOA na neki način estimirati lokacija govornika u vidu kartezijevih (*x*, *y*), odnosno polarnih (*r*,  $\varphi$ ). Ispostavlja se da svaki par mikrofona s pripadajućim TDOA mjerenjem definira jednu hiperbolu kao potencijalnu lokaciju govornika. Dakle, ako imamo više mikrofona onda iz njihovih TDOA mjerenja možemo odrediti lokaciju govornika kao sjecište pripadajućih hiperbola te k tomu još možemo povećati pouzdanost zbog redundantih mjerenja prilikom različitih uparivanja mikrofona. Traženje lokacije izvora pomoću sjecišta hiperbola naziva se hiperbolička multilateracija.

Ovo poglavlje je organizirano kako slijedi: u poglavlju 5.1 opisano je na koji način se iz TDOA mjerenja dobiva hiperbola kao položaj govornika. U poglavlju 5.2 argumentirano je što se točno traži od hiperboličkog algoritma za estimaciju pozicije te kakva rješenja mora dati. U poglavlju 5.3 opisan je Bucherom algoritam hiperboličke multilateracije i dani su rezultati Monte Carlo simulacije dotične metode.

#### 5.1 Hiperbola kao položaj govornika

Položaj para mikrofona i izvora zvuka prikazan je na slici 5.1. Neka su s  $(x_1, y_1)$  te  $(x_2, y_2)$  definirane koordinate mikrofona, a s (x, y) položaj izvora zvuka (govornika). Ne smanjujući općenitost može se pretpostaviti da je izvor koordinatnog sustava na središnjici između mikrofona te da su iznosi koordinata na ordinati mikrofona jednake nuli. S  $R_1$  i  $R_2$  su



Slika 5.1: Položaj izvora zvuka definiran hiperbolom

označene udaljenosti izvora od mikrofona 1 i 2. Udaljenost izvora od mikrofona može se izračunati iz sljedećih relacija:

$$\sqrt{(x-x_1)^2 + y^2} = R_1, \tag{5.1}$$

$$\sqrt{(x-x_2)^2 + y^2} = R_2. \tag{5.2}$$

Ako se s  $c = |x_1| = |x_2|$  označi udaljenost mikrofona od središta koordinatnog sustava te se oduzme (5.2) od (5.1), dobiva se sljedeći izraz:

$$\sqrt{(x-c)^2 + y^2} - \sqrt{(x+c)^2 + y^2} = R_1 - R_2 = R_{12}.$$
 (5.3)

Prebacivanjem s desne strane jednakosti i kvadriranjem cijelog izraza dobije se sljedeće:

$$x \cdot c + \frac{R_{12}^2}{4} = -\frac{R_{12}}{2}\sqrt{(x+c)^2 + y^2},$$
(5.4)

te ponovnim kvadriranjem cijelog izraza dobiva se sljedeća jednakost:

$$x^{2} \cdot (c^{2} - \frac{R_{12}^{2}}{4}) - y^{2} \cdot \frac{R_{12}^{2}}{4} = \frac{R_{12}^{2}}{4}(c^{2} - \frac{R_{12}^{2}}{4}).$$
(5.5)

Ako se uvedu sljedeće oznake:  $R_{12} = 2a$  i  $b^2 = c^2 - a^2$  i izraz (5.5) se podijeli s  $a^2b^2$ , dobiva se klasična jednadžba hiperbole:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, (5.6)$$

koja izražena s početnim varijablama ima sljedeći oblik:

$$\frac{x^2}{\left(\frac{R_{12}}{2}\right)^2} - \frac{y^2}{x_1^2 - \left(\frac{R_{12}}{2}\right)^2} = 1.$$
(5.7)

Pitanje još uvijek ostaje na koji način je povezana estimacija TDOA s hiperbolom. Zvuk putuje konstantnom brzinom *c*, udaljenost mikrofona od izvora je *R* te je vrijeme koje je potrebno da zvuk stigne do mikrofona t = R/c. Ako se zamijeni razlika udaljenosti mikrofona s razlikom produkta brzine i vremena dobije se sljedeća relacija (koja će se u nastavku rada koristiti kao jednadžba hiperbole):

$$\sqrt{(x-x_1)^2 + y^2} - \sqrt{(x-x_2)^2 + y^2} = R_1 - R_2 = c(t_1 - t_2) = c\tau_{12}, \quad (5.8)$$

odnosno, upravo produkt brzine zvuka i TDOA prvog i drugog mikrofona.

#### 5.2 Hiperbolička estimacija pozicije

Lokalizacija govornika poljem mikrofona na mobilnom robotu je karakteristična iz nekoliko razloga; dimenzije polja mikrofona moraju biti razumno male, kao i dimenzija i broj samih mikrofona te algoritam mora dati eksplicitno jednoznačno rješenje. Jasno je na prvi pogled zašto dimenzije moraju biti male, međutim malo više pažnje treba posvetiti eksplicitnom i jednoznačnom rješenju:

- ako se radi lokalizacija u realnom vremenu potrebno je da algoritam da brzo rješenje te iterativni numerički postupci u ovom slučaju se izbjegavaju (osim toga postoji i mogućnost divergiranja rješenja),
- jednoznačno rješenje je potrebno iz razloga što se govornik može nalaziti bilo gdje u rasponu od 0° do 360° oko mobilnog robota te višeznačnosti poput ± rješenja na bilo kojoj osi nisu dozvoljene.

Postoji nekoliko metoda koje pronalaze lokaciju izvora kao sjecište hiperbola. Neke od njih, poput Fangove [13], metode predložene u [3] i Chanove metode [5], daju dvoznačno eksplicitno rješenje i upotrebljavaju se kod lokalizacije u konferencijskim sobama gdje su mikrofoni postavljeni na zidovima. Jedina metoda koja daje eksplicitno, točno i jedinstveno rješenje jest Bucherova metoda predložena u [4].

Bucherova metoda se temelji na algebarskim manipulacijama četiri jednadžbe oblika (5.8), čija TDOA mjerenja su dobivena iz četiri senzora. U dotičnom radu metoda je predložena za GPS lokalizaciju pomoću četiri satelita, a u našem slučaju upotrijebljena su četiri mikrofona. Na koji način sjecište hiperbola definira položaj izvora zvuka prikazano je na slici 5.2.



Slika 5.2: Položaj izvora zvuka kao sjecište hiperbola

# 5.3 Lokalizacija govornika metodom sjecišta hiperbola

Algoritam u suštini koristi redundantna mjerenja da bi se riješio nelinearnosti u jednadžbi oblika (5.8). Izvod algoritma može se naći u [4], no u njemu su korištena 4 senzora da bi se izračunala pozicija u kartezijevim koordinata oblika (x, y,  $\pm z$ ). Višeznačnost po z osi nije problem jer se zna da se izvor nalazi na površini Zemlje. U ovome radu uzet je isti princip, ali je lokalizacija rađena u dvije dimenzije zbog čega je i moguće dobiti
jednoznačno rješenje. Ukratko, četiri mikrofona daju četiri trojke mikrofona od kojih svaka definira jedan pravac. Ako tri mikrofona označimo s *i*, *j* i *k* jednadžba pravca ima sljedeći oblik:

$$y = \frac{R_{ik}x_{ji} - R_{ij}x_{ki}}{R_{ij}y_{ki} - R_{ik}y_{ji}}x + \frac{1}{2}\frac{R_{ik}(R_{ij}^2 + x_i^2 - x_j^2 + y_i^2 - y_j^2) - R_{ij}(R_{ik}^2 + x_i^2 - x_k^2 + y_i^2 - y_k^2)}{R_{ij}y_{ki} - R_{ik}y_{ji}}$$
(5.9)

Dakle, svaka različita trojka daje jedan pravac. Međutim, kako se vidi iz jednadžbe (5.9) jedna trojka može na tri različita načina dati isti pravac. Npr., trojka mikrofona  $\{1, 2, 3\} = \{i, j, k\}$  daje jedan pravac. U jednadžbi (5.9) članovi trojke se uparuje na način  $\{ij, ik\}$ , što se može ostvariti na tri načina:  $\{12, 13\}$ ,  $\{21, 23\}$  i  $\{31, 32\}$ . Dakle svaka trojka mikrofona daje jedan pravac na tri različita načina iz tri različite kombinacije mjerenja, što je bitno iskoristiti jer će se na taj način povećati robusnost. Ako svaka trojka daje po tri pravca, a ima ukupno četiri trojke, krajnji broj pravaca čija sjecišta će se razmatrati je 12. Odnosno, postojati će 12 različitih jednadžbi oblika (5.9).

Da bi se osigurala maksimalna robusnost, u traženju sjecišta trebati će kombinirati svaki način dobivanja jednog pravca sa svim ostalim načinima dobivanja pravca. Tako na kraju dobivamo 54 točke kao potencijalan položaj govornika te se njihova srednja vrijednost uzima kao rezultat jednog koraka algoritma.

Lokalizacija govornika Bucherovim algoritmom testirana je Monte Carlo simulacijama. Prije nego što su TDOA veličine predane algoritmu, zašumljene su s Gaussovim šumom standardne devijacije  $\sigma = 6 \cdot \frac{1}{F_s}$ , jer je to najveća pogreška u uzorcima dobivena u eksperimentima s dva mikrofona. Urađeno je 1000 iteracija s frekvencijom otipkavanja  $F_s = 48$ kHz. Rezultat simulacija za Y geometriju polja mikrofona (čiji će izbor biti argumentiran u poglavlju 6.2) prikazan je na slici 5.3.

Iako Bucherova metoda u slučaju kada mjerenja nisu zašumljena daje točan rezultat, nažalost čim postoji određena nesigurnost metoda podbacuje. Razlog tome je velika mjerna nesigurnost izvan polja mikrofona, jer je i rezolucija manja (o tome više u poglavlju 6.2). Iako se ova metoda ne može primijeniti za lokalizaciju govornika izvan polja mikrofona, simulacije za položaj govornika unutar polja mikrofona su odlične. Tek pri standardnoj devijaciji od  $\sigma = 25 \cdot \frac{1}{F_s}$  (slika 5.4) primijeti se znatnije rasipanje estimata. Histogrami rasipanja estimata po *x* i *y* osima prikazani su na slikama 5.5 i 5.6. Ovo nas navodi na zaključak da bi se dotični algoritam mogao upotrijebiti u inteligentnim prostorima za lokalizaciju



Slika 5.3: Monte Carlo simulacija lokalizacije Bucherovom metodom

govornika i u lokalizaciji mobilnog robota ultrazvučnim senzorima, gdje bi se mikrofoni, odnosno senzori, nalazili u prostoriji. Metoda koja je pokazala dobre rezultate u lokalizaciji teorijom hiperbola objašnjena je u poglavlju 6.1.



Slika 5.4: Simulacija s položajem izvora unutar polja mikrofona



Slika 5.5: Histogram estimata po x-osi



Slika 5.6: Histogram estimata po y-osi

### Poglavlje 6

## Algoritam lokalizacije i praćenja govornika

Pošto se metoda traženja položaja Bucherovom metodom za potrebe ovog diplomskog rada nije pokazala dobrim rješenjem, bilo je potrebno pronaći metodu koja je robusnija na povećanje varijance u mjerenjima. Razlog lošeg ponašanja Bucherove metode jest u lošoj rezoluciji van polja mikrofona, tj., velike udaljenosti između jedne točke gdje se četiri hiperbole sijeku do neke druge. Također nedostatak Bucherove metode jest potreba za četiri TDOA mjerenja radi dobivanja rezultata. Kako će se vidjeti u nastavku, odabrana metoda zahtjeva tri TDOA mjerenja i pokazuje bolje rezultate od Bucherove u Monte Carlo simulacijama.

Ovo poglavlje je organizirano kako slijedi: u poglavlju 6.1 predstavljena je metoda lokalizacije govornika sjecištem asimptota hiperbola, tj., pseudo-linearni estimator pozicije. U poglavlju 6.2 opisana je kvadratna i Y geometrija polja mikrofona te su uspoređene rezolucije dotičnih geometrija. Također je obrazloženo zašto je odabrana Y geometrija mikrofona. U poglavlju 6.3 objašnjeno je na koji način se odabire kombinacija TDOA mjerenja koja predstavljaju položaj govornika. U poglavlju 6.4 dani su rezultati Monte Carlo simulacija algoritma. U poglavlju 6.5 dan je algoritam praćenja govornika koji se temelji na kinematičkom modelu s konstantnim ubrzanjem i RLS estimaciji.

### 6.1 Lokalizacija govornika metodom sjecišta asimptota hiperbola

Jednadžba hiperbole oblika (5.8) se znatno pojednostavljuje ako se zapiše u koordinatnom sustavu koji je s kartezijevim koordinatama povezan na sljedeći način [11]:

$$x = \frac{d}{2} \cosh \xi \cos \varphi,$$
  

$$y = \frac{d}{2} \sinh \xi \sin \varphi,$$
(6.1)

gdje je  $\frac{d}{2}$  konstantan skalirajući parametar.

Zapišimo TDOA jednadžbe (5.8) u (6.1) koordinatnom sustavu. Prvo ćemo razmotriti kartezijev koordinatni sustav (x', y') u kojem se dva mikrofona na koordinatama ( $x_i, y_i$ ) i ( $x_j, y_j$ ) nalaze na koordinatama  $x' = \pm \frac{d}{2}$  i y' = 0, u kojem slučaju:

$$c\tau_{ij} = \sqrt{(x' - d/2)^2 - {y'}^2} - \sqrt{(x' + d/2)^2 - {y'}^2} = d\cos\varphi, \qquad (6.2)$$

što je jednako rezultatu dobivenom pod (4.2).

Azimut, koji ovaj put se mjeri od središnjice spojnice para mikrofona, jednak je sljedećem izrazu:

$$\theta = \alpha + \varphi, \tag{6.3}$$

gdje je  $\alpha$  kut koji koordinatna os x' zatvara s x-osi, odakle slijedi:

$$\alpha = \operatorname{atan2}\left(\frac{y_j - y_i}{x_j - x_i}\right). \tag{6.4}$$

Treba primijetiti da se kutevi  $\alpha$  i  $\theta$  mjere od *x*-osi, odnosno kada su jednaki nula poklapaju se s njom, kao što se i kut  $\varphi$  kada je jednak nuli poklapa s *x*'-osi; kutevi su pozitivni u smjeru suprotnom od kazaljke na satu.

Jednadžbe (6.2), (6.3) i (6.4) omogućuju nam da odredimo azimute  $\theta_{ij}^{\pm}$  asimptota hiperbola para mikrofona *i* i *j*:

$$\theta_{ij}^{\pm} = \operatorname{atan2}\left(\frac{y_j - y_i}{x_j - x_i}\right) \pm \operatorname{arccos}\left(\frac{c\Delta\tau_{ij}}{\|(x_j, y_j) - (x_i, y_i)\|}\right).$$
(6.5)

Dakle, cijela ideja algoritma je aproksimirati hiperbole njihovim asimptotama i tražiti njihovo sjecište (sjecište pravaca), umjesto sjecišta hiperbola. Također je još važno napomenuti da ovdje se isto uzima far-field pretpostavka, jer što je govornik udaljeniji od polja mikrofona, to asimptota bolje opisuje hiperbole i estimacija je točnija.

#### 6.1. LOKALIZACIJA GOVORNIKA METODOM SJECIŠTA ASIMPTOTA HIPERBOLA

Računajući azimut preko (6.5), iznos kuta se sada kreće od 0° do 180° u smjeru obrnutom od kazaljke na satu, gledajući od središnjice spojnice para mikrofona.

Svaka hiperbola daje par asimptota definiranih s (6.5). Za *N* mikrofona dobiva se  $2 \cdot \binom{N}{2}$  asimptota, od kojih  $\binom{N}{2}$  pokazuje u smjeru govornika. Na koji način se odabiru asimptote objašnjeno je u poglavlju 6.3. Sljedeće što treba učiniti, jest odrediti položaj govornika iz dobivenih asimptota.

Središnjice spojnica parova mikrofona *i* i *j* definirane su na sljedeći način:

$$\mathbf{m}_{ij} = \frac{1}{2} \begin{bmatrix} x_i + x_j \\ y_i + y_j \end{bmatrix}_{2\mathbf{x}\mathbf{1}}.$$
(6.6)

67

Da bi se odredio položaj govornika koristi se pseudo-linearni estimator (PLE) temeljen na metodi najmanjih kvadrata (LS) predložen u [11]. PLE lokacije *r* govornika dan je sljedećim izrazom:

$$\hat{r}_{\text{PLE}} = (\mathbf{A}^{\mathrm{T}} \mathbf{A})^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{b}, \qquad (6.7)$$

gdje je

$$\mathbf{A} = \begin{bmatrix} \sin \varphi_{12} & -\cos \varphi_{12} \\ \vdots & \vdots \\ \sin \varphi_{ij} & -\cos \varphi_{ij} \end{bmatrix}_{\binom{N}{2} \times 2}, \qquad (6.8)$$
$$\mathbf{b} = \begin{bmatrix} \begin{bmatrix} \sin \varphi_{12} & -\cos \varphi_{12} \\ \vdots \\ \\ \sin \varphi_{ij} & -\cos \varphi_{ij} \end{bmatrix}_{1 \times 2} \mathbf{m}_{ij} \end{bmatrix}_{\binom{N}{2} \times 1}.$$

Ovdje { $\mathbf{m}_{ij}, \varphi_{ij}$ } predstavlja sve moguće azimute mikrofona *i* i *j*, gdje je  $\varphi_{ij} \in \{\theta_{ij}^-, \theta_{ij}^+\}$ .

PLE se temelji na određivanju azimuta para mikrofona i kao takav još uvijek ima višeznačnost koja se očituje ± znakom u jednadžbi (6.5).

Važno je ovdje spomenuti i drugačije pristupe lokalizaciji govornika koje se mogu naći u znanstvenim radovima. Tako u [18] estimira se položaj koristeći slučajne konačne skupove, u [16] problem nelinearnosti jednadžbe (5.8) i praćenje riješeno je proširenim Kalmanovim filtrom (EKF), a u [21], [20] i [22] koriste se čestični filtri za praćenje govornika čije se kretanje temelji na Langevinovom slučajnom gibanju. Ova rješenja karakterizira stohastički pristup problemu, ali svi oni za temelj imaju TDOA estimaciju te je stoga u ovome radu velika pozornost posvećenja potonjem. Također svi radovi osim [20] rađeni su za lokalizaciju unutar konferencijskih soba, gdje problem višeznačnosti rješenja je lako rješiv.

#### 6.2 Geometrija polja mikrofona

Kao što je konstatirano u poglavlju 4.11, pogreška mjerenja azimuta povećava se kako se približava kutevima 0° ili 180° para mikrofona te postoji problem višeznačnosti iznosa kuta (da li se govornik nalazi ispred ili iza para mikrofona). Radeći s više od dva mikrofona rješava problem višeznačnosti, a upotrebljavajući specifičnu geometriju polja mikrofona ublažava problem povećane pogreške.

Razmotriti će se dvije geometrije polja mikrofona: kvadratna i Y geometrije (slika 6.1 prikazuje spomenute geometrije).



Slika 6.1: Polja mikrofona različite geometrije

Polja imaju duljine spojnica  $L_s$ , za kvadratnu, i  $L_y$  za Y, geometriju. Svaka geometrija ima ukupno šest različitih mjerenja TDOA te pod pretpostavkom da se sva koriste ukupna duljina spojnica je suma svih mogućih spojnica unutar geometrije. Stoga kvadratna geometrija i Y geometrija imaju sljedeće ukupne duljine spojnica:

$$B_{s} = L_{s}(4 + 2\sqrt{2}),$$
  

$$B_{y} = 3L_{y}(1 + \frac{1}{\sqrt{2}}).$$
(6.9)

Odabrana je Y geometrija polja mikrofona umjesto kvadratne (mikrofoni u vrhovima kvadrata), jer Y geometrija postavlja mikrofone tako da nijedna spojnica para mikrofona nije paralelna s drugima. Imajući spojnice s puno različitih usmjerenja povećava vjerojatnost da će nadolazeći val biti pod  $\varphi = 90^{\circ}$ . Također, Y polje (slika 6.2) ima veću rezoluciju od kvadratnog (slika 6.3) zbog gušćeg sjecišta hiperbola



Slika 6.2: Rezolucija polja mikrofona Y geometrije

#### 6.3 Diskriminacija estimacije azimuta

Kao što se može vidjeti iz poglavlja 6.1, svaki azimut  $\varphi_{ij}$  može imati dvije različite vrijednosti. Koju i kako odabrati biti će objašnjeno u nastavku. Na slici 6.4 može se vidjeti da postoji ukupno 12 različitih ćelija s drugačijim kombinacijama mogućih azimuta. Mikrofoni su upareni



Slika 6.3: Rezolucija polja mikrofona kvadratne geometrije

na način da pozitivne vrijednosti kuta  $\varphi_{ij}$  pokazuju van polja mikrofona (kako su točno mikrofoni upareni vidi se sa slike 6.4). Na primjer, ako se govornik nalazi u ćeliji 1, tada bi upotrijebili azimute { $\theta_{12}^+, \theta_{13}^+, \theta_{14}^+, \theta_{34}^+, \theta_{23}^-, \theta_{24}^-$ } u (6.7) za računanje položaja govornika *r*. Koju kombinaciju iz ćelija upotrijebiti odlučuje se na sljedeći način: prvo se izračuna jednadžba (6.7) za svaku od ćelija, ono rješenje *r* koje ima najveću euklidsku mjeru (tj. udaljenost od središta koordinatnog sustava) je položaj govornika. Sva ostala mjerenja padaju unutar polja mikrofona kao što se može vidjeti sa slike 6.5.

Ovdje je bitno primijetiti, što se vidi sa slike 6.5, da PLE je pristran estimator po udaljenosti, ali ne i po azimutu. Pristranost estimatora po udaljenosti može se riješiti instrumentalnim varijablama (IV, slika 6.6) kako je predloženo u [7]. U tom slučaju estimator je konzistentan i asimptotski nepristran. Zašto to nije od interesa u ovome radu biti će obrazloženo u poglavlju 6.4.



Slika 6.4: Podjela po ćelijama polja mikrofona

#### 6.4 TDOA estimacija s N mikrofona

Radeći s N mikrofona moguće je izračunati  $\binom{N}{2}$  različitih kroskorelacija, od kojih je N – 1 nezavisno. Kako pretpostavljamo far-field slučaj i zato što se planarna valna fronta giba konstantnom brzinom, za TDOA vrijednosti vrijedi sljedeći izraz:

$$\Delta \tau_{ij} = \Delta \tau_{kj} - \Delta \tau_{ki}. \tag{6.10}$$

Na primjer, ako radimo s 4 mikrofona, postojati će 3 nezavisna TDOA mjerenja, dok se ostala 3 mogu izračunati iz (6.10). U algoritmu samo oni blokovi podataka koji zadovoljavaju 12 jednadžbi oblika  $\Delta \tau_{ij} = \Delta \tau_{kj} - \Delta \tau_{ki} < \delta$  se smatraju dobrima i nadalje se obrađuju ( $\delta$  je parametar empirijski postavljen kao cjelobrojni višekratnik perioda uzorkovanja  $\frac{1}{F_{i}}$ ).

Odlučivši se o algoritmu TDOA estimacije, geometriji polja mikrofona i načinu hiperboličke lokalizacije, sljedeći logičan korak je testiranje sustava kroz niz simulacija.

Testiranje je izvršeno Monte Carlo simulacija. Tri mikrofona su postavljena u vrhove jednakostraničnog trokuta (duljine stranice a = 0.6 m, a četvrti u ortocentar trokuta. Položaj izvora je u (x, y) = (1.5, 1.5) m, svako mjerenje je zašumljeno Gaussovim šumom standardne devijacije  $\sigma = 6 \cdot \frac{1}{F_s}$ ,



Slika 6.5: PL estimacija svih ćelija

gdje je frekvencija otipkavanja bila  $F_s = 48$  kHz. Slika 6.7 pokazuje rezultate simulacije. Nažalost, i u ovom slučaju estimacija udaljenosti ima preveliku mjernu nesigurnost, ali zato se estimacija po azimutu pokazala vrlo ohrabrujuća. Histogram estimatora po azimutu s pripadajućom srednjom vrijednosti i varijancom, prikazan je na slici 6.8.

Svi TDOA algoritmi estimacije mjere vrijeme kašnjenja temeljeno na diskretnim uzorcima signala. TDOA je stoga cjelobrojni višekratnik perioda otipkavanja. Rezolucija po kutu je određena s koliko različitih DOA mjerenja se može napraviti između 0° i 180°. Pretpostavljajući da je razmak između senzora *d*, brzina zvuka *c* te frekvencija otipkavanja *F*<sub>s</sub>, lako se može pokazati da je maksimum TDOA koji se može izmjeriti  $dF_s/c$ , a minimalni  $-dF_s/c$ . Stoga, kako se može vidjeti iz jednadžbe 4.2, broj različitih DOA mjerenja ovisi o gustoći intervala [ $-dF_s/c$ ,  $dF_s/c$ ]. Pošto na brzinu zvuka u zraku ne možemo utjecati ostaje samo promjena frekvencija otipkavanja i razmaka mikrofona. Da bi se povećala preciznost estimacije TDOA, može se na "umjetni" način povećati frekvencija otipkavanja (jer povećanje dimenzija polja nije rješenje). Naime, povećanje



Slika 6.6: PLE položaja govornika s instrumentalnim varijablama

frekvencije otipkavanja vrši se interpolacijom uzoraka za određeni faktor. Tako je u simulacijama povećana frekvencija otipkavanja 5 puta da bi se simulirao efekt interpolacije. Kako se obrada signala u algoritmu estimacije TDOA vrši u frekvencijskoj domeni, najprikladnije je interpolaciju također raditi u frekvencijskoj domeni. Ovaj postupak radi povećanja preciznosti estimacije TDOA predložen je u [9].

#### 6.5 Algoritam praćenja govornika

Pretpostavimo da imamo estimate položaja govornika u L vremenskih instanci  $t_k = kT$ , k = 0, 1, ..., L - 1, gdje je T period otipkavanja. Neka  $\hat{s}_k$  označava stacionarni estimat lokacije u vremenskom trenutku  $t_k$ . Pretpostavlja se da je  $\hat{s}_k$  dobiven PL metodom estimacije. Estimacija položaja može se "izgladiti" upotrebljavajući model gibanja koji pretpostavlja poznati početni položaj i brzinu, te konstantno ubrzanje govornika. Treba napomenuti da nam praćenje brzine i ubrzanja nije od



Slika 6.7: Monte Carlo simulacija PLE-a s Y geometrijom mikrofona

interesa, već se samo koristi radi boljeg ponašanja algoritma. Praćenje je ostvareno sljedećim kinematičkim modelom, koji predstavlja model s konstantnim ubrzanjem [8]:

$$\mathbf{s}_k = \mathbf{s}_0 + \mathbf{v}_0 t_k + \frac{1}{2} \mathbf{a} t_k^2 = \mathbf{M}_k \boldsymbol{\xi}, \qquad (6.11)$$

gdje su  $\mathbf{s}_0$  i  $\mathbf{v}_0$  položaj i brzina govornika u  $t_0$ , **a** je konstantno ubrzanje, a  $\mathbf{M}_k$  2x6 matrica:

$$\mathbf{M}_{k} = \begin{bmatrix} 1 & 0 & t_{k} & 0 & \frac{1}{2}t_{k}^{2} & 0\\ 0 & 1 & 0 & t_{k} & 0 & \frac{1}{2}t_{k}^{2} \end{bmatrix},$$
(6.12)

i  $\xi$  je 6x1 vektor gibanja:

$$\boldsymbol{\xi} = \begin{bmatrix} \mathbf{s}_0 \\ \mathbf{v}_0 \\ \mathbf{a} \end{bmatrix}. \tag{6.13}$$

S  $L \ge 3$  stacionarnih estimacija položaja  $\hat{s}_k$ , nepoznati vektor  $\xi$  može se estimirati iz sljedećeg izraza:



Slika 6.8: Histogram estimacije azimuta PL metodom

$$\begin{bmatrix} \mathbf{M}_{0} \\ \mathbf{M}_{1} \\ \vdots \\ \mathbf{M}_{K-1} \end{bmatrix} \boldsymbol{\xi} \approx \begin{bmatrix} \hat{\mathbf{s}}_{0} \\ \hat{\mathbf{s}}_{1} \\ \hat{\mathbf{s}}_{K-1} \end{bmatrix}.$$
(6.14)

Da bi se pratio govornik (6.14) može se riješiti pomoću algoritma rekurzivnih najmanjih kvadrata (RLS):

$$\hat{\boldsymbol{\xi}}_{k} = \boldsymbol{\Phi}_{k}^{-1} \boldsymbol{\phi}_{k'} \tag{6.15}$$

gdje je:

$$\mathbf{\Phi}_k = \lambda \mathbf{\Phi}_{k-1} + \mathbf{M}_k^{\mathrm{T}} \mathbf{M}_k, \quad k = 0, 1, \dots$$
(6.16)

$$\boldsymbol{\phi}_{k} = \lambda \boldsymbol{\phi}_{k-1} + \mathbf{M}_{k}^{\mathrm{T}} \mathbf{\hat{s}}_{k}, \quad k = 0, 1, \dots$$
(6.17)

te je 0 <  $\lambda$  < 1 eksponencijalni faktor zaboravljanja. Dobro je primijetiti da  $\mathbf{\Phi}_k^{-1}$  se može izračunati unaprijed pošto je izraz deterministički i nezavisan o estimacijama položaja  $\hat{s}_k$ . Krajnji RLS estimat položaj dan je s:

$$\hat{\mathbf{s}}_{k}^{\text{RLS}} = \mathbf{M}_{k} \hat{\boldsymbol{\xi}}_{k}.$$
(6.18)

Postavljajući  $\lambda$  malim, frekvencija osvježavanja algoritma može se povećati na uštrb varijance. Konačan azimut se računa iz koordinata dobivenih s (6.18). Pokazano je u [8] da RLS algoritam praćenja pokazuje jednako dobro ponašanje, pa čak nekad i bolje, od nekih sofisticiranijih algoritama poput Kalmanovog filtra. Ideja za budući rad bi bila usporedba navedenih algoritama praćenja i čestičnih filtara. Imajući kompletan algoritam gotov, sljedeći korak je bio testiranje u realnim uvjetima na mobilnom robotu.

### 6.6 Eskperimentalni rezultati konačnog algoritma

Polje mikrofona upotrijebljeno u eksperimentima sastojalo se od 4 svesmjerna mikrofona postavljenih u Y geometriju. Tri su bila postavljena na vrhove jednakostraničnog trokuta (duljine stranica a = 0.6 m), a četvrti u ortocentru. Polje mikrofona bilo je postavljeno na Pioneer 3DX robota kako je prikazano na slici 6.9. Audio sučelje se sastojalo od mikrofona, pretpojačala i eksterne zvučne kartice Maya44 USB (cijela oprema cijene ~ 150 eura). Kako se može vidjeti sa slike 6.9, oprema zadovoljava zahtjeve koji su bili opisani u poglavlju 5.2; malih je dimenzija, lagano i cijeli sklop se napaja iz 12V baterije. Komunikacija zvučne kartice i računala bila je preko USB 1.0 protokola. Frekvencija otipkavanja je postavljena na  $F_s = 48$  kHz, preciznost 16 bita, duljina blokova L = 1024 bita, te je upotrijebljen pristup s kvadratnim prozorom i proširivanjem blokova s nulama. Cjelokupni odziv sustava bio je 21.3 ms.

Napravljena su dva eksperimenta. U prvome govornik je hodao oko robota radeći puni krug i pri tome govoreći "Test jedan, dva, tri" kontinuirano. Rezultati ovog eksperimenta pokazani su na slici 6.10, s koje se može vidjeti da algoritam uspješno lokalizira i prati govornika. PL estimati azimuta imaju nekoliko outliera, ali algoritam za praćenje ih uspješno anulira. Za vrijednost eksponencijalnog faktora zaboravljanja izabrano je  $\lambda = 0.8$ , pošto je empirijski ta vrijednost pokazala najbolje rezultate.



Slika 6.9: Polje mikrofona postavljeno na robotu

U drugom eksperimentu, govornika je izgovarao "Test jedan, dva, tri" krećući se od 0° do 45°, te je potom zašutio mijenjajući kut, te nastavio ponavljati rečenicu krećući se od 315° do 270° približno. Ideja ovoga eksperimenta je bila vidjeti odziv sustava na skokovitu promjenu azimuta. Ova situacija je i realnija za sustav od prethodne, jer je više za očekivati da bi govornici mogli imati periode tišine krećući se oko mobilnog robota. Kako se može vidjeti sa slike 6.11 i u ovom slučaju algoritam uspješno vrši lokalizaciju i eliminira outliere. RLS estimacija ima blago nadvišenje zbog nagle promijene kuta, no ovaj problem se može riješiti tako da algoritam uslijed nagle promijene kuta čeka još nekoliko sljedećih vrijednosti. Naravno, nije cilj da robot cijelo vrijeme naglo i skokovito mijenja svoju orijentaciju. Sva snimanja urađena su u prostorijama Zavoda za automatiku i računalno inženjerstvo.



Slika 6.10: Estimacija azimuta (govornik radi puni krug oko robota)



Slika 6.11: Estimacije azimuta (govornik radi skok po azimutu)

### Poglavlje 7

## Zaključak

U ovome radu razvijeno je i izvedeno audio sučelje i algoritam za otkrivanje i praćenje govornika u okolini mobilnog robota. Audio sučelje se zasniva na mikrofonima, pretpojačalima i eksternoj zvučnoj kartici malih dimenzija i energetskih zahtjeva. Takvo izvedeno lako se može uklopiti s ostalim percepcijskim senzorima mobilnog robota. Sam algoritam za otkrivanje i praćenje govornika temelji se na mjerenjima razlike u vremenu dolaska signala između mikrofona (engl. Time Difference of Arrival, TDOA). Ova mjerenja su temelj svakog lokalizacijskog algoritma te je bilo od velike važnosti izvesti robustan i pouzdan algoritam. Mjesta za napredak ima, kao što je bilo konstatirano u poglavlju 4. Kako svaki par mikrofona s pripadajućim TDOA mjerenjem definira hiperbolu, da bi se izbjegla nelinearnost inherentna u jednadžbi hiperbole, one se aproksimiraju svojim asimptotama te njihovo sjecište definira položaj izvora zvuka. Algoritam za praćenje zasniva se na kinematičkom modelu gibanja sa stalnim ubrzanjem i metodom rekurzivnih najmanjih kvadrata. U svakom koraku algoritma estimira se položaj govornika te se iz koordinata računa azimut (mjerenja udaljenosti imaju preveliku mjernu nesigurnost). Sustav svake 21.3 ms daje estimaciju azimuta.

U nastavku rada se osim poboljšanja algoritma estimacije TDOA, planira implementirati i usporediti algoritme za praćenje stohastičkog pristupa te izvesti metodu koja će biti u mogućnosti pratiti više od jednog govornika. Šira slika cjelokupnog sustava ima u planu fuziju različitih senzora (poput praćenja gibajućih objekata laserskim senzorom, vizualna lokalizacija putem kamere...)

## Bibliografija

- R.J. Baken. *Clinical Measurements of Speech and Voice*. Taylor and Francis Ltd., 1987. [cited at p. 11]
- [2] M. Brandstein and D. Ward. *Microphone Arrays: Signal Processing Techniques and Applications*. Springer, 2001. [cited at p. 16]
- [3] Michael S. Brandstein, John E. Adcock, and Harvey F. Silverman. A closedform location estimator for use with room environment microphone arrays. *IEEE Trans. Speech Audio Proc.*, 5:45–50, January 1997. [cited at p. 59]
- [4] Ralph Bucher and D. Misra. A synthesizable vhdl model of the exact solution for three-dimensional hyperbolic poistioning system. *VLSI Design*, 15 (2):507–520, 2002. [cited at p. 60]
- [5] Y.T. Chan and K.C. Ho. A simple and efficient estimator for hyperbolic location. *IEEE Trans. on Signal Processing*, 42 (8):1905–1915, August 1994. [cited at p. 59]
- [6] Jingdong Chen, Jacob Benesty, and Yiteng (Arden) Huang. Time delay estimation in room acoustic environments: An overview. EURASIP Journal on Applied Signal Processing, 2006:1–19, 2006. [cited at p. 16]
- [7] Kutluyil Dogancay. Passive emitter localization using weighted instrumental variables. *Signal Processing*, 84:487–497, 2004. [cited at p. 70]
- [8] Kutluyil Dogancay and Ahmad Hashemi-Sakhtsari. Target tracking by time difference of arrival using recursive smoothing. *Signal Processing*, 85:667– 679, 2005. [cited at p. 74, 76]
- [9] Kutluyil Dogancay and A. Rahim Leyman. Uwb precision geolocation using fft interpolation. *Proc. on 5th EURASIP Conference*, 2005. [cited at p. 73]
- [10] Kevin D. Donohue, Jens Hannemann, and Henry G. Deitz. Performance of phase transform for detecting sound sources wwith microphone arrays in

reverberant an noisy environments. *Signal Processing*, 85:1677–1691, January 2007. [cited at p. 20]

- [11] Samuel P. Drake and Kutluyil Dogancay. Geolocation by time difference of arrival using hyperbolic asymptotes. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 361–364, 2004. [cited at p. 66, 67]
- [12] Y. Ephraim and D. Malah. Speech enhancement using minimum meansquare error short-time spectral amplitude estimator. *IEEE Trans. on Acoustics, Speech and Signal Processing*, ASSP-32(6):1109–1121, 1984. [cited at p. 21]
- [13] B.T. Fang. Simple solutions for hyperbolic and related position fixes. *IEEE Trans. Aerosp. Electron. Syst.*, 26:748–753, 1990. [cited at p. 59]
- [14] J.W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill Book Company, 1968. [cited at p. 10]
- [15] Wolfgang Herbordt. Sound Capture for Human / Machine Interfaces. Springer, 2005. [cited at p. 9]
- [16] Ulrich Klee, Tobias Gehrig, and John McDonough. Kalman filters for time delay of arrival-based source localization. *InterSpeech*, 2005. [cited at p. 67]
- [17] Charles H. Knapp and G. Clifford Carter. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-24(4), August 1976. [cited at p. 19, 20]
- [18] Wing-Kin Ma, Ba-Ngu Vo, S. Singh, and A. Baddeley. Tracking an unknown time-varying number of speakers using tdoa measurements: A random finite set approach. *IEEE Trans. on Signal Processing*, 54(9):3291–3304, September 2006. [cited at p. 67]
- [19] Kazuhiro Nakadai, Hiroshi G. Okuno, and Hirokai Kitano. Real-time sound source localization and separation for robot audition. *Proceedings of 2002 International Conference on Spoken Language Processing*, pages 193–196, 2002. [cited at p. 56]
- [20] J.-M. Valin, F. Michaud, and J. Rouat. Robust 3-d localization and tracking of sound sources using beamforing and particle filtering. *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 841–844, 2006. [cited at p. 21, 67, 68]
- [21] J. Vermaak and A. Blake. Nonlinear filtering for speaker tracking in noisy and reverberant environments. *ICASSP*, 2001. [cited at p. 67]

- [22] Darren B. Ward, Eric A. Lehman, and Robert C. Williamson. Particle filtering algorithms for tracking an acoustic source in a reverberant environment. *IEEE Trans. on Speech and Audio Processing*, 11(6), 2003. [cited at p. 67]
- [23] Kevin William Wilson. Estimating Uncertanty Models for Speech Source Localization in Real-World Environments. PhD thesis, Massachusetts Institute of Technology, 2006. [cited at p. 20]

# Popis slika

2.1	Veza između kartezijevih i sfernih koordinata	8
2.2	Širenje ravnog i sfernog vala iz točkastog izvora	11
2.3	Skica impulsnog odziva prostorije	13
3.1	Kroskorelacija sa i bez PHAT težinske funkcije	21
3.2	Blok dijagram algoritma estimacije TDOA	22
0.2		
4.1	Računanje DOA iz TDOA	24
4.2	Položaji govornika tokom eksperimenta	26
4.3	Pucketanje prstima	26
4.4	Izgovor "Test"	27
4.5	Izgovor "Test jedan, dva, tri, test jedan, dva, tri"	27
4.6	GCC i GCC-PHAT pri pucketanju prstima na 0°	28
4.7	GCC i GCC-PHAT pri izgovoru "Test" na 0°	29
4.8	GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na 0°	30
4.9	Praćenje pri izgovoru "Test jedan, dva, tri" na 0°	31
4.10	Histogram na $0^{\circ}$	31
4.11	GCC i GCC-PHAT pri pucketanju prstima na $-15^{\circ}$	32
4.12	GCC i GCC-PHAT pri izgovoru "Test" na –15°	33
4.13	GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na $-15^{\circ}$ .	34
4.14	Praćenje pri izgovoru "Test jedan, dva, tri" na $-15^{\circ}$	35
4.15	Histogram na $-15^{\circ}$	35
4.16	GCC i GCC-PHAT pri pucketanju prstima na $-30^{\circ}$	36
4.17	GCC i GCC-PHAT pri izgovoru "Test" na –30°	37
4.18	GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na $-30^{\circ}$ .	38
4.19	Praćenje pri izgovoru "Test jedan, dva, tri" na $-30^{\circ}$	39
4.20	Histogram na $-30^{\circ}$	39
4.21	GCC i GCC-PHAT pri pucketanju prstima na $-45^{\circ}$	40

4.22	GCC i GCC-PHAT pri izgovoru "Test" na $-45^{\circ}$	41
4.23	GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na $-45^{\circ}$ .	42
4.24	Praćenje pri izgovoru "Test jedan, dva, tri" na $-45^{\circ}$	43
4.25	Histogram na $-45^{\circ}$	43
4.26	GCC i GCC-PHAT pri pucketanju prstima na $-60^{\circ}$	44
4.27	GCC i GCC-PHAT pri izgovoru "Test" na $-60^{\circ}$	45
4.28	GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na $-60^{\circ}$ .	46
4.29	Praćenje pri izgovoru "Test jedan, dva, tri" na $-60^{\circ}$	47
4.30	Histogram na $-60^{\circ}$	47
4.31	GCC i GCC-PHAT pri pucketanju prstima na 70°	48
4.32	GCC i GCC-PHAT pri izgovoru "Test" na 70°	49
4.33	GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na 70° .	50
4.34	Praćenje pri izgovoru "Test jedan, dva, tri" na 70°	51
4.35	Histogram na 70°	51
4.36	GCC i GCC-PHAT pri pucketanju prstima na $-80^{\circ}$	52
4.37	GCC i GCC-PHAT pri izgovoru "Test" na –80°	53
4.38	GCC i GCC-PHAT pri izgovoru "Test jedan, dva, tri" na –80°.	54
4.39	Praćenje pri izgovoru "Test jedan, dva, tri" na –80°	55
4.40	Histogram na $-80^{\circ}$	55
E 1	Deležaj izvora zvuka definiran hinerholem	EO
5.1	Položaj izvora zvuka kao ciocičto hiperbolo	50
5.Z	Monto Carlo cimulacija lokalizacije Bucherovom metodom	60
5.5 E 4	Simulacija o položejem izvore uputer polje mikrofono	62
5.4 5.5	Simulacija s polozajem izvora unutar polja mikrotona	63
5.5 E.C		04
5.6	Histogram estimata po y-osi	64
6.1	Polja mikrofona različite geometrije	68
6.2	Rezolucija polja mikrofona Y geometrije	69
6.3	Rezolucija polja mikrofona kvadratne geometrije	70
6.4	Podjela po ćelijama polja mikrofona	71
6.5	PL estimacija svih ćelija	72
6.6	PLE položaja govornika s instrumentalnim varijablama	73
6.7	Monte Carlo simulacija PLE-a s Y geometrijom mikrofona	74
6.8	Histogram estimacije azimuta PL metodom	75
6.9	Polje mikrofona postavljeno na robotu	77
6.10	Estimacija azimuta (govornik radi puni krug oko robota)	78
6.11	Estimacije azimuta (govornik radi skok po azimutu)	79

# Popis tablica

4.1	Rezultati eksperimenta pucketanja prstima pri 0°	28
4.2	Rezultati eksperimenta izgovora "Test" pri 0°	29
4.3	Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri $0^\circ$ .	30
4.4	Rezultati eksperimenta pucketanja prstima pri $-15^{\circ}$	32
4.5	Rezultati eksperimenta izgovora "Test" pri –15°	33
4.6	Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri –15°	34
4.7	Rezultati eksperimenta pucketanja pr stima pri $-30^{\circ}$	36
4.8	Rezultati eksperimenta izgovora "Test" pri –30°	37
4.9	Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri –30°	38
4.10	Rezultati eksperimenta pucketanja pr stima pri $-45^{\circ}$	40
4.11	Rezultati eksperimenta izgovora "Test" pri $-45^{\circ}$	41
4.12	Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri –45°	42
4.13	Rezultati eksperimenta pucketanja pr stima pri $-60^{\circ}$	44
4.14	Rezultati eksperimenta izgovora "Test" pri $-60^{\circ}$	45
4.15	Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri –60°	46
4.16	Rezultati eksperimenta pucketanja pr stima pri $70^\circ$	48
4.17	Rezultati eksperimenta izgovora "Test" pri 70°	49
4.18	Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri 70° .	50
4.19	Rezultati eksperimenta pucketanja pr stima pri $-80^{\circ}$	52
4.20	Rezultati eksperimenta izgovora "Test" pri –80°	53
4.21	Rezultati eksperimenta izgovora "Test jedan, dva, tri" pri –80°	54

## Abstract

In this undergraduate thesis, a method for speaker localization and tracking is proposed based on Time Difference of Arrival (TDOA) estimation. Firstly, TDOA estimation is enhanced with so called PHAT- $\beta$  (Phase Transform) function and tested with two microphone array. The localization method is based on Pseudo-linear estimator (PLE), and Y-shaped array for spatial sampling is proposed and compared to square array. The tracking is realized with Recursive Least-Squares algorithm (RLS). At the end, results recorded on a mobile robot are presented, showing that the developed audio interface and algorithm can localize and track speaker in mobile robot environment.

**Key words:** azimuth estimation, TDOA estimation, DOA estimation, hyperbolic multilateration, speaker localization, microphone array.

# Životopis

Ivan Marković rođen je 17. travnja 1985. u Osijeku, s prebivalištem u Belom Manastiru. Godine 1999. upisuje III. gimnaziju u Osijeku. Za treći razred srednje škole dobiva stipendiju pod pokroviteljstvom američke organizacije ASSIST te provodi godinu dana pohađajući Verde Valley School u Sedoni, Arizona, SAD. Godine 2002. završava treći razred s odličnim uspjehom te se vraća u Hrvatsku završiti srednješkolsko obrazovanje, koje završava 2003. godine s odličnim uspjehom. Tokom srednje škole sudjelovao je redovito na natjecanjima iz matematike te u četvrtom razredu i iz filozofije. Godine 2002. postaje viceprvak Hrvatske u karateu te dotičnim rezultatom postaje član Hrvatske karate reprezentacije. Godine 2003. upisuje Fakultet Elektrotehnike i Računarstva. Od druge godine studija, zbog ostvarenog prosjeka ocjena, prima državnu stipendiju. Od nagrada tokom studija dobio je: za odličan uspjeh na trećoj godini nagradu "Josip Lončar", na petoj godini za izvrsnost postignutu tokom studija nagradu "INETEC" Instituta za nuklearnu tehnologiju, te je iste godine bio finalist Top stipendije za top studente. Trenutačno je apsolvent smjera Automatika.