

Informatika Ingeniaritzako Gradua

Konputazioa

Gradu Amaierako Lana

Robot sozial baten hizkerari laguntzen dion gorputz adierazpenean aurrerapausoak

Egilea

Oihane Cantero Dominguez

2021

Informatika Ingeniaritzako Gradua
Konputazioa

Gradu Amaierako Lana

**Robot sozial baten hizkerari laguntzen dion
gorputz adierazpenean aurrerapausoak**

Egilea

Oihane Cantero Dominguez

Zuzendaria(k)

Elena Lazkano eta Igor Rodriguez

Laburpena

Robotikan bultzada handia hartzen ari den arloetariko bat robotika soziala da. Izan ere, robotak geroz eta gertuago dauzkagunez gizartean, haiekin komunikazio on bat bermatzea ezinbestekoa da. Gizaki eta roboten arteko interakzioa ahalik eta naturalena izateko, interakzio hori bi gizakiren arteko elkarrekintzatik hurbiltzea premiazkoa da. Horretarako, garrantzitsua da robotek gizakiok hitz egiten dugun bezala hitz egitea, hau da, hitz egin ahala beso, buru eta enborrarekin mugimendu koherenteak ere egitea. RSAIT taldeak Pepper robotari solasaldian gure antzeko mugimenduak egiteko ahalmena ematen dion portaera garatu du. Mugimendu horiek, aldiz, ez dute inongo loturarik robotak esaten duenarekin, hizkeraren erritmoari lagundu baino ez diote egiten. Proiektu honetan, termino konkretuekin erlazio semantikoa duten mugimenduak modu koherentean txertatu dira jatorrizko portaera horretan, robotaren gestikulazio ahalmena areagotuz eta naturalagoa bihurtuz.

Gaien aurkibidea

Laburpena	i
Gaien aurkibidea	iii
Irudien aurkibidea	v
Taulen aurkibidea	vii
1 Sarrera	1
1 Sarrera	3
1.1 Testuingurua	3
1.2 Motibazioa	7
1.3 Proiektuaren helburua	8
1.4 Erabilitako tresnak	9
1.4.1 Hardwarea	9
1.4.2 Softwarea	11
1.5 Kudeaketa	13
1.5.1 Bete beharreko atazak	13
1.5.2 Ataza bakoitzareztat estimatutako ordu kopurua	14
1.5.3 Lanaren banaketa denboran zehar	15

2	Garapena	17
2	Literaturaren azterketa	19
2.1	Mugimendu motak	19
2.2	Ahots eta mugimenduaren arteko sinkronizazioa	21
2.3	Solasaldiko mugimenduak sortzeko beste hurbilpen batzuk	22
3	Hizkerari lotutako mugimenduen txertaketa	25
3.1	Mugimenduen adierazpena	26
3.2	Mugimenduen hautaketa	28
3.2.1	Hautaketa estrategia	28
3.2.2	Hautatutako terminoak eta keinuak	29
3.3	Txertatze prozesua: hurbilpen ezberdinak	30
3.3.1	Lehen hurbilpena	30
3.3.2	Bigarren hurbilpena	32
3.4	Mugimenduak txertatzeko probabilitatea	33
3.5	Sinkronizazioa	34
4	Esperimentazioa eta emaitzak	37
3	Ondorioak eta etorkizunerako lana	41
5	Ondorioak eta etorkizuneko lanak	43
5.1	Ondorioak	43
5.2	Etorkizuneko lanak	44
Eranskinak		
A	Esteka interesgarriak	47
Bibliografia		49

Irudien aurkibidea

1.1	Robot itxura ezberdinak	4
1.2	<i>Eccerobot</i> eta <i>roboy</i> robotak	6
1.3	<i>Softbank</i> <i>Roboticseko Pepper</i> eta <i>Nao</i>	7
1.4	<i>Pepper</i> robota eta bere motorrak	10
1.5	<i>Pepper</i> robota <i>Choregraphe</i> ingurunean	12
1.6	Proiektuko <i>Gantt</i> diagrama	16
3.1	Sistemaren arkitektura orokorra	25
3.2	Mugimendu baten adierazpena: adibidez	27
3.3	Hitz gakoei lotutako mugimenduak	29
3.4	Lehen hurbilpenaren eskema	30
3.5	Lehen hurbilpena	31
3.6	Bigarren hurbilpenaren eskema	32
3.7	Bigarren hurbilpena	33

Taulen aurkibidea

1.1	Ataza bakoitzareztat aurreikusitako eta benetan egindako ordu kopurua . . .	14
4.1	Hitz gakoak eta haien pisuak	38
4.2	Bi eta hirugarren probetan txertatutako mugimenduak	39

1. ATALA

Sarrera

1. KAPITULUA

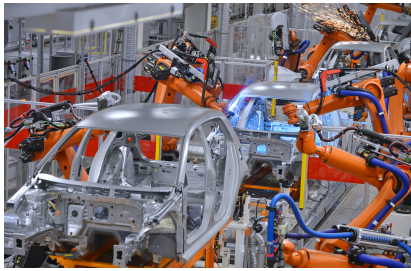
Sarrera

Robot "modernoak" XX. mendearen bigarren erdian garatzen hasi ziren arren, batez ere azken hamarkadetan eman du aurrerakada handia esparruak, adimen artifizialean eta ikasketa sakonaren arloan izan diren aurrerakuntzak medio. Gaur egun, mota askotako robotak daude, giza itxurakoak, animalia itxurakoak, beso mekanikoak, edo roverrak, adibidez, eta gainera, ataza mota ezberdin askotarako erabiltzen dira. Hauek guztien erabilerak eta helburuak oso ezberdinak izan daitezke, adibidez, industrian, irakaskuntzan, medikuntzan, nekazaritzan, erabiltzen dira.

1.1 Testuingurua

Robot zabalduenak eta ezagunenak ingurune industrialetan topatzen dira, eta osagaiak automatikoki eraikitze edo mugitzeko erabiltzen diren beso mekanikoak dira, [1.1a](#) irudian agertzen diren modukoak. Hauetariko gehienak ataza errepikakorrak egiteko erabiltzen dira eta ingurune kontrolatu eta babestuetan jarduten dira, pertsonekin inongo elkarrekin-tzarik eduki gabe, martxan dauden bitartean.

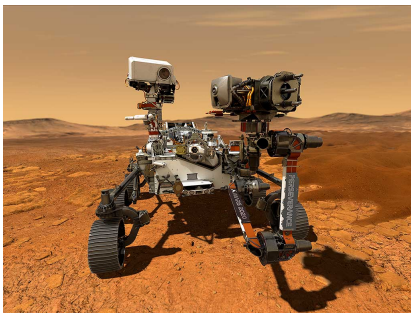
Etixerako robotak, gaur egun, *Roomba* itxurakoak dira gehienak (ikus [1.1b](#)). Nabigazio sistema bat dute, garbitzeko bidea sortzeko eta oztopoak ekiditeko. Gainera, karga estaziora laguntzarik gabe itzultzeko ahalmena ukan behar dute bateria guztia deskargatu baino lehen. Estilo berdinean belarra mozten duten robotak daude, baita igerilekuak garbitzeko balio dutenak ere. Horiek ere nabigazio sistema bat ukan behar dute, baina nabigatzeko



(a) Industriako robotak



(b) Roomba



(c) Perseverance



(d) Sophia



(e) Nuka

1.1 Irudia: Robot itxura ezberdinak

helburua espazioko puntu bat izan ordez, libre dauden puntu guztiak atzitzea dute helburu, hala nola, belardi osoa moztu eta igerilekuaren hondo osoa garbitu.

Bere kasaz mugitzeko gai diren beste robot batzuk, adibidez, beste planetak esploratzeko erabiltzen diren *Rover*ak dira. Horietariko bat 2020ko uztailean bidali eta 2021eko otsailean Martitzera iritsi den *Perseverance* robota da (ikus 1.1c irudia). Bere helburua Martitzen noizbait bizitzarik egon den edo ez ikertzea da, lurreko laginak hartu eta aztertea, eta Martitzeko atmosferatik oxigenoa sortzen saiatzea da, besteak beste. Mugitzeko 6 gurpil ditu eta sentso-re pila bat, haien artean, kamerak, mikrofonoak, radarrak, etab. Gainera helikoptero itxurako drone bat ere darama, hegaldi autonomo baten probak egiteko [Farley et al., 2020].

Baina robot hauen interesa autonomoak izatea da, izan ere, gorputz fisikoa ukan behar dute (ez dira simulatuak izan behar), ingurunearekin elkarreragiteko sentsoak eta eragi-leak behar dituzte eta erabakiak modu autonomoan aritzeko gaitasuna behar du (denbora luzean kanpoko laguntzarik gabe funtzionatzeko gaitasuna). Hori lortuta, gizakien laguntzarik gabe, edo oso laguntza gutxirekin haien lana betetzea lortuko dute.

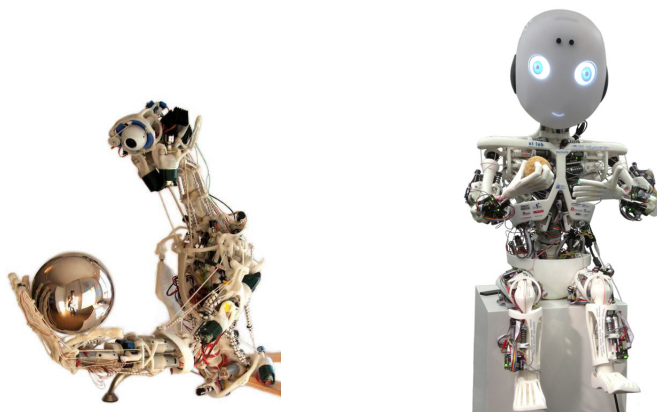
Guzti horien artean, robot sozialak aurkitzen ditugu. Haien helburua gizakiekin interakzioa sortzea da, eta interakzio hori ahalik eta naturalena izatea. Horretarako, robotek ezaugarri konkretu batzuk betetzea lagungarria da, hala nola, autonomia izatea, gizaki edo beste robot sozialekin elkarrizketetan parte hartzea edo helburu zailak betetzeko beste robotekin elkarlanean aritzeko gaitasuna ukaitea [Daily et al., 2017].

Itxura guztietakoak izan daitezke baina gehienak humanoide edo animalia itxurakoak dira, eta erabilera nagusiak dendetan jendea bideratzea, etxean edo irakaskuntzan laguntzea eta osasun arreta ematea dira, besteak beste. Orokorrean jendeari laguntzeko sortuak izan dira.

Animalia itxura dutenen artean, bi adibide *Sonyren Aibo* txakurtxoak edo *Nuka* itsas txakurra (1.1e. irudia) dira. *Nuka* itsas txakurra [Cabibihan et al., 2013] lanean erabili dute, autismoa duten haurrak laguntzeko. Dementzia arazoak dituzten helduekin ere probak eginak izanak dira [Šabanović et al., 2013]. Robot honen ezaugarri nagusia bere itxura da, eta gainera, laztantzerakoan edo harekin hitz egitean erreakzionatzeko gaitasuna du. Pazienteak erlaxatzeko, estresa kentzeko, edo motibatuzeko lagungarria izan daitekeela ikusi dute. Gainera, animalia bizidun batek ukan ditzakeen betebeharririk ez du, ez zaio janaririk eman behar, eta ez du hondakinik sortzen. Hala ere, honek galdera etikoak sortzen ditu, besteak beste, ea robotek animaliak ordezkatzen ahal dituen, edo ea *Nuka* erabiltzeak gaixo horiei kasu gutxiago egitea ekarriko duen, adibidez [Calo et al., 2011].

Beste robot batzuek aldiz, gizaki itxura ahalik eta errealistena izateko helburua dute, adibidez *Hanson Robotics* enpresako *Sophia* robota (1.1d. irudia). Bere helburua ikerketaren, irakaskuntzaren, eta aisialdiaren arloetan aurrerakuntzak egitea da. Sentimenduak erakusteko 62 aurpegi espresio ezberdin egin ditzake bere azala osatzen duen silikonari esker. Aurpegiak ezagutzeko eta galderei erantzuteko gai da, eta bere hizlari ezberdinei egokitzeko gai da (kultura, hitz egiteko modua, emozioak...). Bestalde, herrialde bateko herritartasuna lortu duen lehen robota izan da, Saudi Arabiakoa hain zuzen ere. Horrek gizartean eztabaida batzuk sortu ditu, hala nola, robotikaren etorkizunari buruzko gogoeta, robot adimentsuei lotutako etika, edo gizartean robotek hartuko duten lekuari buruzko gogoetak [Retto, 2017].

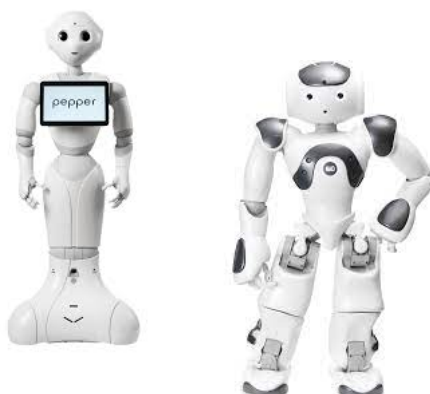
Gizakien antzekoak izaten saiatzen diren beste batzuen artean, kanpoko itxura ordez mugimendu eta artikulazio mailan imitatzen saiatzen direnak daude, horiek robot antropometikoak dira. Gizakien antzeko eskeleto bat dute eta horri esker gure antzeko mugimenduak egin ditzakete [Diamond et al., 2012]. Horien artean ezagunetarikoa *ECCEROBOT* eta *roboy* dira, 1.2 irudian ikusten den bezala. Esan bezala, hauen berezitasuna gizakien zinatika sistema imitatzea da. Hezurak, tendoiak eta muskuluak ordezkatzen dituzte material ezberdinekin, eta horri esker gizakiok egin ditzakegun mugimendu gehienak kopiatzen saiatzen dira, motor elektriko eta engranajeek sortzen dituzten mugimenduen zurruntasuna gainditzeko.



1.2 Irudia: *Eccerobot* eta *roboy* robotak

Ukan ditzaketen erabileren artean, irakaskuntza dago, besteak beste 1.3. irudiko *Softbank Robotics* taldeko *NAO* eta *Pepper* robotak erabiliz. Lan honetan erabili den robota *Pepper* izan da. Hau enpresa, museo edo edozein lekuko harreran erabil daiteke, jendea bideratzeko, laguntzeko zein gomendioak emateko hitzez edo bularraldean duen tabletari esker.

Beste erabilera bat, autismoa duten haurrekin terapia gisa erabiltzea izan daiteke.



1.3 Irudia: *Softbank Roboticeko Pepper eta Nao*

[Cabibihan et al., 2013] artikuluan robot sozial askorekin ataza ezberdinak betetzeko saia-kerak egin dituzte, eta ikusi dute robot ezberdinak terapia ezberdinentzako erabil daitezkeela. Helburua haur horiek laguntzeko egoerak sortzea eta robotarekin interakzioak sortzea da. Erraztasun gehiago ukanen dute robotarekin jolasteko eta erlazioak sortzeko, eta hori haientzat oso lagungarria izan daiteke ondoren haien inguruko jende eta lagunekin harremanak sortzeko.

Bukatzeko, jendearen zaintzan ere erabiltzen dira, ospitaletan eta adineko pertsonen zaintzan. Adinekoekin, egunerokoan kontaktu soziala mantentzeko lagungarriak izan daitezke, eta baita laguntzeko edo haien ongizatea hobetzeko ere. Gainera, telepresentzia egin dezaketen robotak ere garrantzitsuak izan daitezke, lagun, familia edo medikuekin komunikazioa errazteko [Mois and Beer, 2020]. Ospitaletan orohar pertsonen egoera psikologikoa hobetzeko erabiliak izan dira, estresa kentzeko, motibatuzeko edo konpainia emateko [Scoglio et al., 2019].

1.2 Motibazioa

Aurreko atalean aipatu bezala, robot sozialak arlo oso ezberdinetan erabiltzen dira, eta robot sozial horien helburuetariko bat gizaki eta roboten arteko interakzioa eraginkorra eta ahalik eta naturalena izatea bermatzea da. Izan ere, gizaki eta roboten arteko elkarrekintza hobetzeko komunikazioa oso garrantzitsua da, komunikazioa hobetzeak gizakiok dugun jarreratik hurbilduko baititu.

Gizakiok gure artean komunikatzeko hizkera, mugimenduak, aurpegiko espresioak eta intonazioa erabiltzen ditugu, besteak beste. Hitz egitean egiten ditugun mugimenduek

askotan esaten dugunarekin lotura dute eta esan nahi dugunaren ulermenerako laguntzen dute. Hortaz, roboten eta gizakien arteko komunikazioa hobetu nahi badugu, haien hitz eta mugimenduen artean lotura bat egotea gauza ona izan daiteke.

RSAIT taldeak Pepper robotari solasaldian gure antzeko mugimenduak egiteko ahalmena ematen dion portaera garatu du. Mugimendu horiek, aldiz, ez dute inongo loturarik robotak esaten duenarekin, hizkeraren erritmoari lagundu baino ez diote egiten. Proiektu honetan, termino konkretuekin erlazio semantikoa duten mugimenduak modu koherentean txertatu dira jatorrizko portaera horretan, robotaren gestikulazio ahalmena areagotuz eta naturalagoa bihurtuz. Helburua robotak egingo dituen mugimenduen eta esandako hitzen artean lotura bat egotea da, horretarako, hitz gako batzuei mugimendu adierazgarri batzuk lotuko dizkiegu, eta elkarrekin esan eta egin beharko dira. Izan ere, gizakiok hitz egiten dugun bitartean egiten ditugun mugimenduen artean, badaude batzuk besteak baino adierazgarriagoak direnak. Hortaz, proiektu honen ideia mugimendu horiek robotari eginaraztea da.

1.3 Proiektuaren helburua

Proiektuaren helburua robotak hitz edo hitz segida konkretu bat esaten duenean horri lotutako mugimenduak egitea da. Adibidez, norbaiti agur esatean eskuarekin keinu bat egitea. Noski, robotaren portaera naturala gertatzeko, mugimendu horiek noiz eta zein maiztasunarekin txertatu beharko diren erabaki beharko da.

Proiektuaren abiapuntu gisa, aurretik sortutako beste bi tresna daude: *MovementGAN* eta *animations.pml* proiektuak. Lehena, RSAIT taldeak sortutakoa da [Zabala et al., 2019], eta robotarentzat mugimenduak sortzen ditu, zeintzuk gizakiok hitz egiten dugunean egiten ditugunen antza duten. Denbora zehatz bat emanda sarrera gisa, denbora hori irauten duen mugimendu segida logiko eta leun bat ematen du. Bigarrena mugimendu andana bat da, eta bertan emozioen edo mugimendu motaren arabera robotak exekutatzeko prest dauden mugimenduak daude. Hauen artean hautatu beharko da zeintzuk izango diren hitz gakoiei lotu direnak. Hortaz mugimenduen sorrera bera ez da proiektuaren helburua, nahiz eta batzuk aldatzea beharrezkoa izango den, txertatu beharreko lekuan sartzeko luzeegiak baitziren.

Robotak esaten duen testutik abiatuta, mugimenduak sortuko dira eta sortutako mugimendu horien artean, hitz gako batzuei lotutako mugimenduak txertatuko dira, hitz gako horiek esaten diren bitartean. Horretarako, lehenik hitz gako horiek zeintzuk diren defi-

nitu behar da, eta, ondoren, esaldietan hitz horiek dauden bilatu. Baldin badaude, haien posizioa gorde eta, bukatzeko, mugimendua behar den momentuan txertatu beharko da. Gainera, naturalagoa izateko, mugimenduak ez dira beti txertatu beharko, ezta beti mugimendu bera txertatu beharko, gizakiok ere ez baitugu sistematikoki mugimendu bat hitz batekin lotzen. Portaera naturala lortzeko, esaldika mugimendu bakarra txertatua izango da gehienez eta mugimendu hori zein den ausaz hautatuko da.

1.4 Erabilitako tresnak

Atal honetan erabilitako software eta hardwarea zein izan den azaltzen da.

1.4.1 Hardwarea

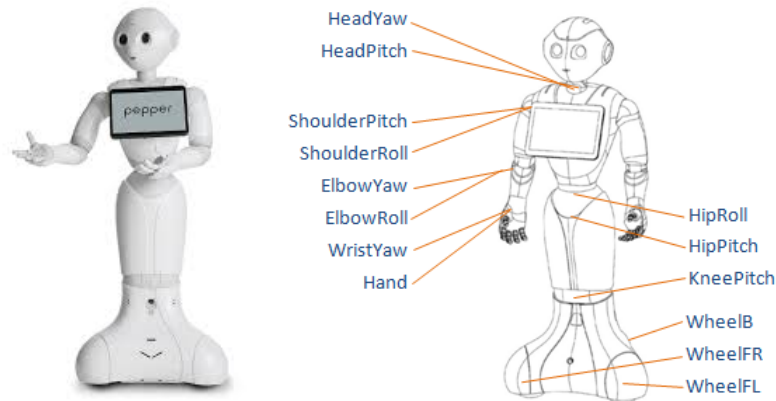
Erabiliko den hardware nagusia *Pepper* robota bera izango da, eta inplementatutako lanaren probak egiteko erabilia izango da.

Pepper robota

Garatutako kodearen probak zuzenean 1.4. irudian ikusten den *Pepper* robotean probatuko dira. *Pepper* robotak 120cm neurtzen ditu, 28kg pisatzen ditu eta bularraldean komunikazioa laguntzen duen pantaila bat du. Aurpegiak eta gizakien oinarrizko emozioak desberdintzeko gai da.

Ingurunearekin elkarreragiteko sentsoare eta eragile hauek ditu:

- Bi bozgorailu buruaren ezker eta eskuinean.
- Zortzi LED begi bakoitzaren inguruan, hamar bozgorailu bakoitzaren inguruan ("belarrien" inguruan) eta bat sorbalda bakoitzaren gainean.
- Lau mikrofono buru gainean.
- Hiru kamera begien inguruan (bi 2D kamera eta 3D sentsoare bat).
- *IMU (Inertial Movement Unit)* bat, giroskopio eta azelerometro batez osatua robotaren hiru ardatzekiko errotazio-aldaketa eta ardatzekiko gertatzen den azelerazioa kalkulatzeko. *



1.4 Irudia: *Pepper* robota eta bere motorrak

- Sei laser. *
- Bi sentsore infragorri. *
- Lau sonar (ultrasoinu sentsoreak), bi sentsore eta bi eragile bi kausuetan bat robotarekiko aurrera begira eta bat atzera begira. *
- 30 kodegailu magnetiko.
- Hogei motor mugitzeko. 1.4. irudian ikus daitezkeenak.
- Bost sentsore taktil, hiru buruan eta bat esku bakoitzean.
- Hiru talka sentsore, gurpil bakoitzaren ondoan bat.

*IMUa, laserrak, sentsore infragorriak eta sonarrak 1.4. irudiko *KneePitchen* kokatuak dira, belauen mailan.

Gainera, bularraldean pantaila bat du, komunikatu bitartean ulermena laguntzeko, adibidez, norabide bat erakusteko edo emozio batzuk adierazteko erabil daiteke. Baina lan honetan ez da erabiliko, robotaren mugimenduetan zentratu garelako.

Elementu desberdinak mugitzeko, 1.5. irudian ikusten den bezala, hogeitaz askatasun gradu ditu: bi buruan, bi sorbalda bakoitzean, bi ukondo bakoitzean, bi eskumutur bakoitzean, bat esku bakoitzean, bi gerrian, bat belauetan eta hiru oinarrian duen mugimendu sistema holonomoa gauzatzeko. Hortaz, hogeitaz askatasun gradu horiek hamazazpi giltzadura (*joints*) eta nabigaziorako hiru gurpildun mugimendu sistema daude. Gorputz espresiorako erabiltzen diren mugimenduak egiteko, hamazazpi giltzadurak erabiliko dira [Pandey and Gelin, 2018].

Robota kontrolatzeko, wifi bidez konektatzeko aukera dago, *IEEE 802.11 a/b/g/n* erabiliz. Hori erabili da egindako lana robotean probatzeko. Bateriak 8 ordu behar ditu osoki kargatzeko eta, erabileraren arabera, 7 eta 20 ordu arteko autonomia du.

Mahaigaineko PCa

Kodea garatzeko eta robotarekin komunikatzeko Ubuntu 20.04 sistema eragilea duen ordenadore eramangarria erabili da. Honek wifi txartela behar du, robotarekin konektatu ahal izateko. Gainera, bertan, beharrezko software guztia instalatu beharko da garapenarekin hasi aurretik.

1.4.2 Softwarea

Robota simulatzeko *Choregraphe* programazio grafikorako tresna erabili da, *Aldebaranem* robotak (*Pepper*, *NAO*, *Romeo*) simulatzeko tresna da, eta inplementatzeko garaian eta probak egiteko ezinbesteko tresna izan da. Bestalde, robota kontrolatu eta simulazioak egin ahal izateko *NAOqi SDK* konfiguratu behar izan da, bertan erabili diren *API*ak daukelako. Gainera, robotak esaldiak esan ahal izateko, hau da, testutik audioa sortzeko, *Text To Speech* (TTS) tresna ezberdinak erabili dira.

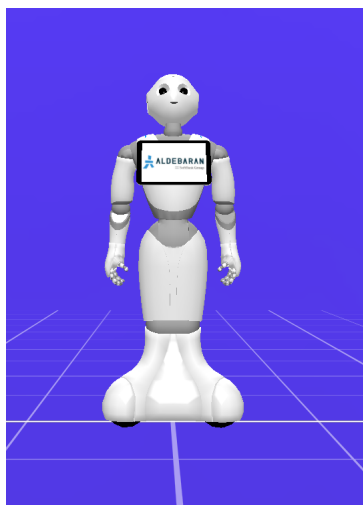
Choregraphe

Choregraphe Pepper robota simulatzeko softwarea da. Robotaren mugimenduak sortzeko edo aldatzeko aukera ematen du, eta mugimenduak proiektuan txertatu baino lehen simulazioan proba daitezke. Gainera, robota kontrolatu daiteke, eta *Python* lengoaiari esker hobekuntzak egin daitezke ere bai. 1.5. irudian *Pepper Choregraphen* nola agertzen den ikus daiteke.

Simulatuz sortutako kodearen probak egiteaz gain, txertatu nahi ditugun mugimenduak *animations.pml* proiektutik hautatzeko eta aldatzeko erabili da.

NAOqi SDK

Robota kontrolatu ahal izateko *NAOqiren* SDK erabili behar da. Horrek *NAOqi Frameworka* erabiltzea ahalbidetzen du robota simulatzeko, harekin komunikatzeko eta informazioa partekatzeko. Robotaren mugimenduak, audioa eta tabletan agertzen diren bideoak



1.5 Irudia: Pepper robota *Choregraphe* ingurunean

koordinatzeko erabil daiteke. *Python* lengoia erabiliz, ordenagailuan (*Choregrapher* esker) edo robotean bertan inplementatutako kodea exekutzea ahalbidetzen du. Izan ere, kodea *Choregraphen* probatu ahal izateko ezinbestekoa da robotaren simulazioa abiaraztea, eta hori egiteko SDK hau konfiguratzeko beharrezkoa da.

Text To Speech tresnak

Proiektuko momentu ezberdinetan testua audio bilakatzeko bi tresna ezberdin erabili dira, *Google Text To Speech* (gTTS) eta *IBM Watsoneko Text To Speech*. Izan ere, ez dituzte konfigurazio aukera berdinak eta ez dituzte ahots berdinak sortzen ere ez.

gTTS testua ahots bilakatzeko ahalbidetzen duen *Googleko* tresna bat da. Testuaren hizkuntza zein den jartzeko aukera dago (azentua eta intonazioa aldatzeko batez ere), testua polikiago esateko aukera dauka eta sortutako audioa gordetzeko aukera ematen du.

IBM Watsoneko Text To Speech *IBM* enpresak sortutako tresna bat da. Honek ere testua audioan bilakatzen du, baina ahotsa aldatzeko aukera ematen du. Eta hizkuntza bakoitzeko bizpau ahots ezberdin eskaintzen ditu. Honek ez du funtziorik audioa lokalean gordetzeko, orduan ondoren eskuz gorde behar da.

Python eta pythoneko moduluak

NaoQi-k *Pythonen* 2.7 bertsioan onartzen duenez, hori erabili beharko da proiektuan zehar. Gainera, mugimenduak sortzeko alde aurretik garatuta dagoen portaerak *Tensor-*

flowen 1.10. bertsioa, eta *Kerako* 2.2.4 bertsioa eskatzen ditu. Azkenik, *gTTS* eta *pydub* ere ezinbestekoak dira.

1.5 Kudeaketa

Kudeaketari dagokionez, lehenik proiektua egiteko bete beharreko atazak zeintzuk diren definitu dira. Bakoitza betetzeko denbora bat estimatu da eta ataza gehienak garaiz eginak izan dira.

1.5.1 Bete beharreko atazak

Lanean bete diren atazak hiru taldetan banatu dira: antolaketa edo kudeaketa, literaturaren azterketa, inplementazioa eta txostena eta aurkezpena prestatzea. Zati ezberdin horietan ataza hauek definitu dira:

- Kudeaketa:
 - Bete beharreko atazak definitu.
 - Bakoitza gauzatzeko ordu kopurua estimatu.
- Literaturaren azterketa.
- Inplementazioa:
 - Hitz gakoak hautatu.
 - Hitz gako bakoitzari lotutako mugimenduak zeintzuk izango diren hautatu.
 - Mugimenduak hitz gakoak esateko behar den denborari egokitu.
 - Hitzak testuan aurkitzeko estrategia diseinatu.
 - Mugimenduak non txertatu definitu.
 - Mugimenduak haien lekuan txertatu.
 - Bete mugimendu berdina ez egiteko modu bat definitu.
- Txostena eta aurkezpena:
 - Txostenaren idazketa.

Ataza	Aurreikusitako ordu kopurua	Ordu kopuru erreala
Hitzak hautatu	5	7
Mugimenduak hautatu	10	7
Mugimenduak egokitu	17	20
Hitzak testuan aurkitu	45	40
Mugimendua txertatzeko posizioa aurkitu	45	40
Mugimendua txertatu	55	55
Mugimendu ezberdinak egin	30	25
Atazak definitu	10	10
Literaturaren azterketa	10	25
Txostena idatzi	40	45
Aurkezpena prestatu	25	30
Tutoretzak eta probak robotean	15	15
Guztira	307	319

1.1 Taula: Ataza bakoitzareztat aurreikusitako eta benetan egindako ordu kopurua

- Aurkezpenaren prestaketa.

Ataza hauetariko batzuk aldi berean egin dira eta beste batzuk aldiz, proiektuaren osoan zehar beteak izan dira.

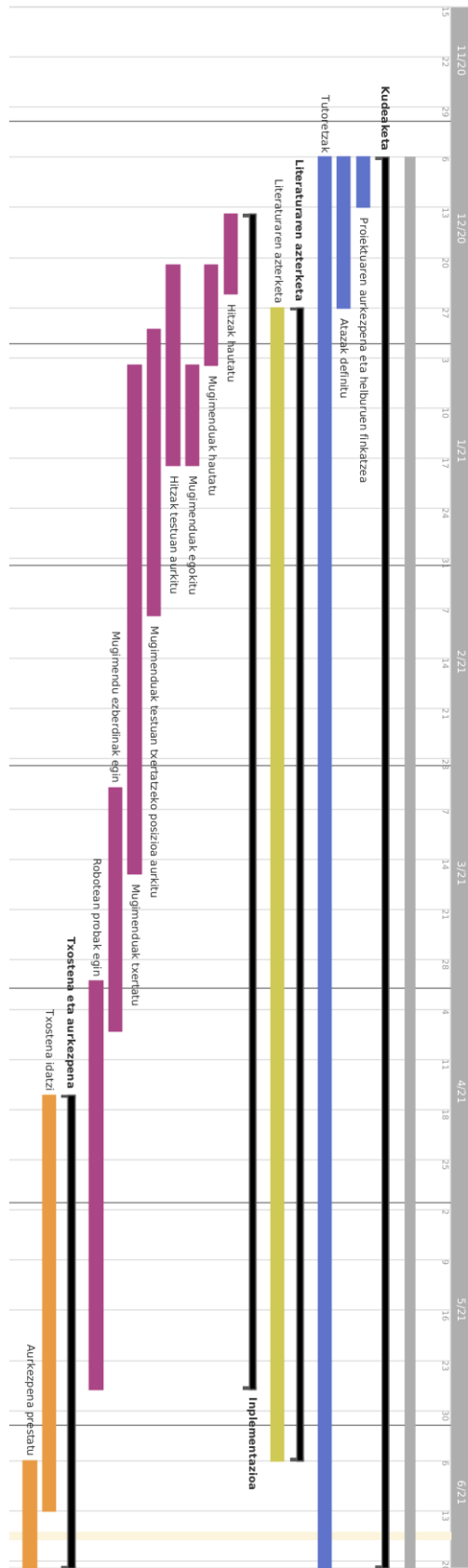
1.5.2 Ataza bakoitzareztat estimatutako ordu kopurua

1.1.taulan ataza bakoitzerako aurreikusi zen denbora eta benetan emandako denborak agertzen dira. Ikus daiteke ataza batzuetan estimatutako denbora baina gehiago eman dela eta beste batzuetan gutxiago baina gutxi gorabehera estimatutako bete da.

Ataza batzuetan estimatutako denbora gainditu da, proiektua hasi baino lehen aurreikusiak izan ez ziren aldaketa batzuk egon direlako. Adibidez, proiektu erdialdean testu aldaketa bat egon da eta hitzen eta mugimenduen hautaketa berriz egin behar izan da. Bestalde, mugimenduen txertaketarako bi hurbilpen landuak izan direnez, ataza batzuk errepikatu behar izan dira.

1.5.3 Lanaren banaketa denboran zehar

[1.6.](#) irudian proiektuko *Gant* diagrama erakusten da. Bertan ataza bakoitza noiz egin den ikus daiteke. Atazak lau taldetan banatuak daude: kudeaketa, literaturaren azterketa, inplementazioa eta txostena eta aurkezpena prestatzea, [1.5.1.](#) atalean azaldu bezala.



1.6 Irudia: Proiektuko Gantt diagrama

2. ATALA

Garapena

2. KAPITULUA

Literaturaren azterketa

Robotika sozialeko arloan ikerketa asko egin da, eta lan honen oinarri gisa hartu direnak bereziki solasaldian egiten ditugun mugimendu ezberdinei buruzkoak (gizaki edo robotentzat), denboran zehar hitzak eta mugimenduak sinkronizatzeari buruzkoak eta sare neuronalei buruzkoak dira. Mugimendu hauek guztiak, gorputz espresioari lotutakoak dira, hemen nabigazioa ez baitugu kontutan hartzen.

Alde batetik, mugimenduen sorrerari buruzkoak daude, gizakiok mugimenduak nola egiten ditugun aztertzen dituztenak, eta hortik mugimendu mota ezberdinak identifikatu dira. Izan ere, mugimenduak nola sortzen ditugun ulertzea ezinbestekoa da robotek errepikatzea nahi bada. Eta bestetik, mugimenduak nola osatuak diren eta hauen sorrera nola egiten den kontutan hartuta, imitatzen saiatzen diren algoritmoak sortu dira.

Hortaz, atal honetan ikerketa mailan literaturan topatu ditugun lanak aztertuko dira. Mugimendu mota ezberdinak, mugimendu eta hitzen arteko sinkronizazioa nola egiten den eta gure antzeko helburua lortzeko topatu diren hurbilpen desberdinak aztertuko dira.

2.1 Mugimendu motak

Hitz egiten dugunean mota askotako mugimenduak egiten ditugu, besoekin egindako mugimenduak, buruarekin egindakoak, edo begirada eta aurpegiarekin egindakoak. Mugimendu horiek esanahi ezberdinak ukan ditzakete, adibidez, gehienetan eskuak deskribatzeko erabiltzen dira eta aurpegia emozioak pasarazteko [Wagner et al., 2014].

Gorputzaren enborrekin eta besoekin egiten diren mugimenduak mugimendu ikertuenak izan dira, horiek baitira esanguratsuenak. Mugimendu horietan bost fase nagusi definitu dira [[Bressem and Ladewig, 2011](#)]:

1. Atsedean fasea: hasierako posizioa, mugimendua hastera doa.
2. Prestaketa fasea: fase honetan mugimendua prestatzen da, hurrengo fasea hasteko prestaketa da.
3. Mugimendu "kolpea": informazio gehiena hemen ematen da, mugimendua bera egiten den fasea da. Hau da mugimenduko informazio gehien ematen duena, eta beraz, faserik garrantzitsuena.
4. Kontrol fasea: mugimendurik ez duen fase bat da, eta mugimendua baino lehen edo ondoren gerta daiteke.
5. Berreskuratze fasea: fase honetan eskuak hasierako posiziora itzultzen dira, atsedean fasean zeuden bezala egoteko.

Gainera, gorputz eta besoen mugimendu horiek mota ezberdinekoak izan daitezke, hala nola gauza bat erakusteko, esaten duguna deskribatzeko edo esaten dugunari indar handiagoa emateko. Horiez gain, mugimendu nabarmen edo ezagun batzuk ere kontutan hartzen dira (hatz lodia gora, etab.), edo zerbait baieztatze edo ezeztatze egiten ditugun mugimendu txikiak [[Thomaz et al., 2016](#), [Wagner et al., 2014](#)]. Mugimendu hauek guztiak orokorrean bost kategoriatan sailkatzen dira [[Kucherenko et al., 2020](#), [Wagner et al., 2014](#)]:

- Keinu enblematikoak: keinu hitzartuak dira, "hatz lodia gora", adibidez. Esanahi konkretu bat dute eta normalean ikusten direnean zuzenean jakin daiteke zer esanahi duten.
- Keinu ikonikoak: inguruko objektu edo mugimendu konkretu bat deskribatzeko erabiltzen dira. Adibidez, esfera edo baloi bat nolakoa den mimatze erabilitako mugimendu bat.
- Keinu metaforikoak: kontzeptu abstraktuak azaltzeko erabiltzen diren keinuak dira hauek. Inguruan ez dagoen objektu bat deskribatzeko erabiltzen dira. Adibidez, begi bistan ez dugun edozein gauzaz hitz egitean egindako keinu bat.

- Keinu deiktikoak: inguruko lekuak erakusteko erabiltzen dira. Adibidez, "hemen" edo "han" esatean eskuarekin erakutsiz egindako keinu bat.
- "Taupada" keinuak: eskuekin egiten diren mugimendu motz eta azkarrak dira, aho-tsa eta mugimenduak koordinatzeko erabiltzen dira eta intonazioa eta ahotsaren ozentasuna segitu ohi dute. Ez dute loturarik esandakoarekin, ez dute esanahi semantikorik.

Buruaren mugimenduei dagokienez, sentimenduak erakusteko, zerbait baieztatu edo ezeztatzeko edo esan nahi duguna indartzeko erabiltzen dira, eta hauek intentsitate edo abiadura ezberdinetakoak izan daitezke [[Wagner et al., 2014](#)].

Buruaren mugimenduez gain, begiradak ere bere garrantzia du, elkarriketa erregulatzeko erabili daitekeelako. Norbaiti zerbait emateko erabilgarria izan daiteke, baita begirada-kontaktua bermatzeko eta konfiantza harreman bat sortzeko ere. Gainera, zerbaiti buruz hitz egitean, seinalatzeko ere oso erabilgarria da [[Thomaz et al., 2016](#)].

Robot batek mugimendu horiek egiteko, autonomia minimo bat izatea ezinbestekoa da. Sheridan eta Verplank ikerlarien arabera, makina baten autonomia maila gizakiak egin behar duen lanaren arabera da (geroz eta autonomia handiagoa, orduan eta gizakiaren laguntza txikiagoa izango da) [[Sheridan and Verplank, 1978](#)].

2.2 Ahots eta mugimenduaren arteko sinkronizazioa

Hitz egitean mugimenduak egiten badira, eta esaten denarekin lotura bat baldin badute, mugimenduak eta hitzak sinkronizatu behar dira, robotaren portaera inkoherentea gerta daiteke eta. Hori oso garrantzitsua da, bestela egiten diren mugimenduek zentzua galtzen dute. Onargarria izan ohi da mugimenduak hitzak baino lehenago egiten hastea, baina ulergarria izateko bien arteko denbora tarte 600ms-koa baino txikiagoa izan behar da. Sinkronizazioa egokia gertatzeko hiru erregela proposatuak izan dira:

1. Sinkronizazio fonologikoaren erregela: mugimenduaren hasiera silaba garrantzitsuena baino lehen agertu beharko da.
2. Sinkronizazio semantikoaren erregela: elkarrekin gertatzen diren mugimendu eta hitzek ideia berdina irudikatu behar dute.

3. Sinkronizazio pragmatikoaren erregela: elkarrekin gertatzen diren mugimendu eta hitzek funtzio pragmatiko berdina dute.

Lehen erregela hori mugimendu deiktikoekin (gauzak erakusteko mugimenduak) askotan gertatzen dela konturatu dira. Hau da, erakusteko mugimendua hitza baino lehen egikaritzen hasten dira. Gainera, mugimendu konplikatuak egiteko, mugimendua "konprimitu" egiten da eta, batzuetan, silaba garrantzitsuena esan baino lehen bukatzen dugu mugimenduaren zati garrantzitsuena egiten.

Hortaz, 600ms-ko denbora tarte onargarri hori esplikatzen modu bat hitzak sortzea mugimenduak sortzea baina "zailagoa" dela izan daiteke. Izan ere, elkarrizketa batean esandako hitzen eta egindako mugimenduen arteko lotura eta bakoitzaren sorrerari buruzko hipotesia ezberdinak daude. Orohar, hipotesia guztiak adosten dira mugimenduak sortzea esaldi zuzenak sortzea baina azkarragoa dela, mugimenduak sortzean hizkuntzaren prozesamendurik egiteko beharrik ez dagoelako eta mugimenduak azkarrago sortzen direnez, lehenago hasten dira gertatzen. Hortaz, batzuek diote lehenik mugimenduak sortuko liratekeela eta horien gainean hitzak egingo liratekeela. Beste hipotesia batzuek aldiz, biak paraleloan egiten direla diote, eta hitzak sortzea mugimenduak sortzea baina luzeagoa denez, hitz egiten mugimenduak egiten baina lehenago hasiko ginateke [Wagner et al., 2014].

2.3 Solasaldiko mugimenduak sortzeko beste hurbilpen batzuk

Mugimendu motak eta sinkronizazioa kontutan hartuz, solasaldiko mugimenduak sortzeko zenbait aurrekari topa ditzakegu. Izan ere, gizaki eta roboten arteko elkarrekintzan lan andana bat egin da, 90. hamarkadatik aurrera batez ere [Goodrich and Schultz, 2007].

Alde batetik, beste tresna batzuk erabiliz [Le et al., 2011] artikuluan NAO robota eta simulagailu bat erabili dituzte. Mugimendu adierazgarriz osatutako bi datubase dituzte (bat robotarentzat eta beste bat simulagailuarentzat, ezberdinak direlako) eta hortik testuinguruaren arabera diskurtsoari lotutako mugimenduak egiten dituzte. Mugimendu batzuen sorrera arazoak ukan dituzte, simulagailua eta robota ez baitziren berdinak, horregatik bi datubase ezberdin sortu behar izan dituzte.

Elkarrizketa osoari dagokionez, [Stiefelhagen et al., 2004] artikuluan, gizaki eta roboten arteko elkarrekintza osoa kontutan hartu dute, robota gizakiaren postura analizatzeko gai

da, hau da, nora begiratzen duen prozesatzeko gai da, eta egiten dituen keinu deiktikoak ulertzeko gai da. Probak egiteko sukalde baten eszena sortu dute, robotari tresna batzuk piztu, itzali edo hartzeko eskatzen zaio eta aginduak ulertu, bete eta erantzuteko gai da. Hortaz, robota gizakiarekin elkarrizketa bat ukaiteko gai da, eta hitz egiten duen bitartean mugimenduak egiten ditu, baina hauetariko gehienak gauzak erakusteko mugimenduak dira (keinu deiktikoak). Mugimendu horiek sortzeko, lehenik gizakiak esan diona prozesatzen du, eta horren arabera eskatutakoa betetzeko argibide gehiago behar baditu galderak egingo ditu, eta bestela behar den ekintza egingo du. Mugimendu eta beharrezkoak diren hitz guztiak (kasu honetan, sukaldeko tresnen izenak) datubase batean gordeta daude.

Azken urte hauetan sare neuronalen arloa oso azkar garatzen ari da, eta horren ondorioz robotika sozialean eta roboten hizkera laguntzeko mugimenduen sorreran ikasketa sakonari loturiko berrikuntzak egon dira. Sare neuronalak giza burmuina imitatzen duten sistemak dira. Ataza askotarako erabiltzen ahal dira eta erabili baino lehen entrenatu behar dira. Neurona geruzaz osatuak dira eta ondoan dauden bi geruzetako neuronen arteko erlazioa adierazteko pisuak erabiltzen dira. Pisu horiek entrenamendu fasean egokitzen dira, ondoren ahalik eta emaitza onenak emateko. Erabilpenen artean, irudien ezagutza, itzulpen automatikoa edo hemen egiten den mugimenduen imitazioa izan daitezke.

Oinarri gisa hartzen den sistema [Zabala et al., 2019]k sortutakoa da. Entrenamendurako datuak jasotzeko, kamara baten bitartez hizkerari laguntzen dieten mugimenduak harrapatu dituzte. Horrela, denbora bat emanda, luzera horretako mugimendu sorta natural bat itzultzen du.

[Kucherenko et al., 2020] artikuluan, hitz egiten ari diren pertsonen bideoz osatutako datubase batetik abiatuz, audioa eta mugimenduak harrapatu dituzte, eta sare neuronala entrenatzeko erabili dituzte. *Feed-forward* motako sare neuronal bat da eta irteera mugimenduak osatzeko artikulazio ezberdinek egin behar dituzten errotazio angeluez osatutako zerrenda bat da. Sortutako mugimenduak hizkerari laguntzen dioten mugimenduak dira, eta batzuek esandako hitzekin lotura dute eta beste batzuek ez.

Emaitzen kalitatea neurtzeko, metrika objektibo eta subjektibok erabili dituzte. Metrika objektiboen artean, sortutako mugimenduen azelerazio aldaketak azertu dituzte besteak beste. Subjektiboei dagokienez, jendeari bideoak konparatzeko eskatu diete, bi bideo elkarren ondoan erreproduzitzen, zeinek duen gizaki itxura handiena esan behar zuten, adibidez. Sistemaren probak simulatzailean egin dituzte baina robot humanoideetan erabiltze-

ko aukera ere dago ¹. Gainera, modeloa aldatuz (elementu batzuk txandaka kenduz), eta sarrera aldatuz emandako emaitzak ere begiratu dituzte, horrela, sistemako zein elementu den garrantzitsua ikusi ahal izan dute [Kucherenko et al., 2020].

Beste lan honetan, [Alexanderson et al., 2020], modu ezberdinetan sortutako mugimenduak konparatu dituzte. Entrenatzeko aurreko artikuluko datubase berdina erabili dute, hau da, mugimenduak eta ahotsa sortzeko modeloak pertsona bakar baten mugimendu eta ahotsen gainean entrenatu dituzte. Gainera, mugimenduak ahotsetik sortzen dituzte, eta konparaketa *Text To Speech* (TTS) tresna batek sortutako ahotsarekin eta gizaki baten ahotsarekin sortutako mugimenduen artean egin dituzte. Ikus daiteke oraindik ezberdintasunak daudela TTS ahotsetik sortutako mugimenduen eta gizakien ahotsetik sortutakoen artean. Hemen ere probak avatar batekin egin dituzte ².

Bestalde, Gradu Amaierako lan proposamen honetan erabiltzen den eta [Zabala et al., 2019] erreferentziaren antzekoak ere eginak izan dira. Audio batetik abiatuz, *GAN* motako sare neuronal bat erabiliz mugimenduak sortzen dituzte eta ondoren *Pepper* eta *NAO* robotean probatu dituzte. Sarea entrenatzeko *TED* bideoak erabili dituzte, bertan pertsona ezberdin askoren hitz eta mugimenduak baitaude [YU and Tapus, 2020, Yoon et al., 2019]. Bestalde, sarrera moduan audioa baino gehiago kontutan hartzen duten sareak daude. Izan ere, entrenatzean bideoetatik mugimenduak baina baita audioa, esandako testua eta nor hitz egiten duen kontutan hartzen dute. Horrela, gizaki bakoitzak mugimendu ezberdinak egin ditzakegula kontutan har dezakete, eta pertsona ezberdinek mugimendu ezberdinak egi-ten badituzte, nahastea ekidin dezakete [Yoon et al., 2020]. Azken artikulua honetan ere, probak simulatutako avatar batekin egin dituzte.

¹<https://svito-zar.github.io/gesticulator/>

²<https://simonalexanderson.github.io/IVA2020/>

menu orokorrak emango ditu GAN sareak. Bestetik, testu zatian zehar txertatu nahi den mugimendu bati lotutako hitzik baldin badago, mugimenduaren txertaketa egin behar da.

Txertatze prozesu hori egin ahal izateko, lehenik hitz gakoak zeintzuk izango diren definitu eta horiek dagozkien mugimendu adierazgarriak hautatu beharko dira. Ondoren testuan mugimendu horiei lotutako hitzak egonez gero, txertatzeko maiztasuna garrantzitsua izango da. Txertatze prozesua bera da konplexutasun handiena duen urratsa, non txeratu eta GANek sortutako mugimendurik ezabatu behar den aukeratu beharko da eta.

Txertatze puntua definitzeko bi estrategia erabili dira: lehen, eskuz txertatu behar den lekua aurkitzea, eta bigarrena *IBMko Automatic Speech Recognizer (ASR)* eta *Text To Speech (TTS)* tresnak erabiliz aurkitzea. Bi tresna hauek audioan hitz gakoak aurkitzeko eta testua audio bihurtzeko erabiltzen dira hurrenez hurren.

Hortaz, atal honetan lehenik mugimenduak nola adierazten diren azalduko da, eta gero mugimenduen hautaketa, hitzak testuan nola bilatzen diren eta mugimenduen txertaketa nola egiten den azalduko da.

3.1 Mugimenduen adierazpena

Mugimenduak errepresentatzeko hiru ezaugarri hartu behar dira kontutan, zein artikulazio mugituko den, noiz eta nola. Hortaz, mugimendu bakoitza hiru bektorez osatua dago, *names*, *keys* eta *times* deituak. Bakoitzak 14 elementu ditu, eta bakoitzak giltzadura (edo *joint*) bati egiten dio erreferentzia.

Names zerrenda giltzadura edo *joint* bakoitzaren izenaz osatua da. Horietarik 14 daude eta haien kokapena gizakiok ditugun artikulazioetatik hurbiltzen dira, eta 1.4. irudian ikusten da zehazki non dauden kokatuak. *Keys* zerrendan *names* zerrendako elementu bakoitzari doakion giltzadurak egin behar duen mugimenduaren angelua gordetzen da, radianetan. Angelu hauek absolutuak dira, mugimendua egiteko momentuan dagoen angelua ez da kontutan hartzen. Eta *times* zerrendak *keys* zerrendako mugimendu bakoitza zein momentutan exekutatu behar den adierazten du, segundotan.

3.2 irudian mugimendu bat nola osatuta dagoen ikus daiteke. Argiki ikusten dira hiru zerrendak koadroen bidez azpimarratuta, eta bakoitzaren balioak. *Names* zerrendako elementuen ordena garrantzitsua da, izan ere, hiru zerrenden luzera berdina da (hemen 14, 14 *joints* ditugulako) eta hiru zerrendetako posizio berdinean dagoen informazioa elkarrekin doa. Hemengo adibidean, *names* zerrendako 4. indizean '*LHand*' izena baldin badago,

keys zerrendan [0.736364,0.708387] angeluen zerrenda eta *timesen* [1.48,2.08], ezkerreko eskuarekin 0.736364 radianeko mugimendua egingo du 1.48. segundoan eta 0.708287 radianekoa 2.08. segundoan.

```
(['HeadPitch', 'HeadYaw', 'LElbowRoll', 'LElbowYaw',
'LHand', 'LShoulderPitch', 'LShoulderRoll', 'LWristYaw',
'RElbowRoll', 'RElbowYaw', 'RHand', 'RShoulderPitch',
'RShoulderRoll', 'RWristYaw'],
[[[0.234049, -0.612719, -0.537554], [-0.366667, -0.073674,
-0.0782759], [-1.28698, -1.08756, -1.00626], [-1.213,
-1.64142, -1.69051], [0.736364, 0.708387], [0.526121,
0.716335, 0.874338], [0.191986, 0.0889301, 0.161028],
[-0.994838, -0.998676], [1.37604, 1.10759, 1.01708],
[1.44192, 1.78093, 1.78706], [0.736364, 0.708751],
[0.392746, 0.644321, 0.786985], [-0.164061, -0.107422,
-0.162646], [0.846485, 0.831386]],
[[[0.8, 1.56, 2.16], [0.8, 1.56, 2.16], [0.72, 1.48, 2.08],
[0.72, 1.48, 2.08], [1.48, 2.08], [0.72, 1.48, 2.08],
[0.72, 1.48, 2.08], [1.48, 2.08], [0.64, 1.4, 2], [0.64,
1.4, 2], [1.4, 2], [0.64, 1.4, 2], [0.64, 1.4, 2], [1.4,
2]] )
```

3.2 Irudia: Mugimendu baten adierazpena: adibidez

Garrantzitsua da *times* zerrendako elementu bakoitzean dauden zenbakiak txikienetik handienera ordenatuak izatea, bestela errore bat ematen du eta robotak ez du mugimendurik eginen. Izan ere, azpizerrenda bakoitzean artikulazio bakoitzak zein mugimendu noiz egin behar duen agertzen da, eta exekuzioan zehar zerrenda ordenean korritzen denez, ezinbestekoa da *times* zerrendako elementuak ordenatuak egotea. Gainera, *keys* zerrendako elementuak dagokien *times* zerrendako elementuen orden berdinean egon behar dira, eginak izango diren ordenean, hain zuzen.

Bestalde, *timeseko* azpizerrenda bateko bi elementu oso hurbil badaude, hau da, bi mugimendu bata bestearekiko oso hurbil badaude, dagokien *keyseko* azpizerrendako elementuak ez dira oso ezberdinak izan behar, bestela mugimendua oso azkarra izango da. Baliteke mugimendu horiek *Choregrapheko* simulagailuan ongi ibiltzea, baina egiazko robotean kontuz ibili behar da mugimendu azkarrekin.

3.2 Mugimenduen hautaketa

Txertatuko diren mugimenduak hautatzeko, besteak beste, testuingurua eta mugimenduen esanahi semantikoa hartu behar dira kontuan. Horrez gain, hitz sorta bati mugimendu berdina esleitzea edo hitz bati mugimendu bat baino gehiago esleitzea erabaki daiteke, betiere testuinguruaren arabera.

3.2.1 Hautaketa estrategia

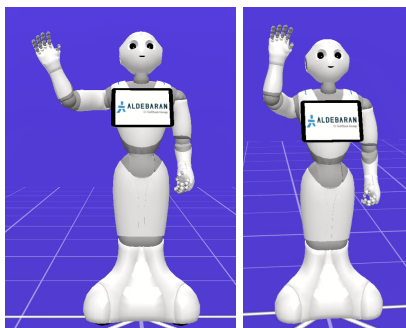
Mugimenduak hautatzeko, adierazgarritasuna hartu behar da kontuan. Orohar mugimenduek ez dute luzeegiak izan behar, bestela mugimendua egiteko behar den denbora hitza esatekoa baino askoz luzeagoa izango da, eta hori ez da naturala eta gainera sinkronizazio arazoak sor daitezke. Kasurik hoberenean hitzak esateko denbora eta mugimenduak egiteko behar den denborak berdina izan behar du, edo gutxienez ahalik eta hurbilen egon behar dute.

Hautatutako termino sorta orokorra izan behar da, hau da, hitzak ahalik eta testu gehiengentan agertu behar dira, testuinguruari oso lotuak egon gabe. Horrela, robotak esango duen testua aldatzen bada, mugimendu eta hitz gakoak ez dira berriz definitu beharko. Gainera, hitz eta mugimendu adierazkorrak hautatu behar dira, izan ere, "eta" hitzari mugimendu bat esleitzen baldin bazaio, informazio gutxi pasaraziko da mugimenduen bidez. "Kaixo" hitza aukeratzen bada, aldiz, informazio gehiago pasarazten dugu, esanahi semantiko handiago dagoelako. Horregatik garrantzitsua da hitz gakoaren zerrenda osatzean keinu bidez adierazten ditugun termino arruntak aukeratzea portaera natural bat lortu nahi badugu. Bestalde, hitz gako bakoitzari mugimendu bat baino gehiago esleitu zaizkio, testuan hitz gako bat aurkitzean beti mugimendu berdina ez egiteko, gizakiok bezala.

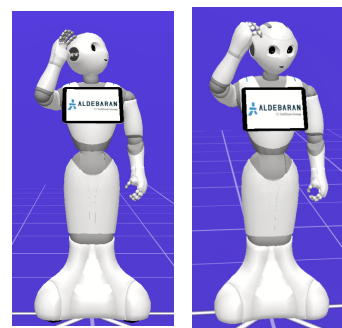
Hortaz, hitz gako ezberdin asko hautatu dira, eta batzuentzat mugimendu posible bat baino gehiago sortu da. Horrela, beti mugimendu berdina ez egitea bermatzen da. Gainera, termino bakoitzarentzat zein maiztasunarekin onartzen den mugimendua egikaritzea ere inportantea da, ez baita interesgarria mugimendua beti egitea. Hau da, batzuetan txertaketa ez da egingo, gizakiok ere ez baititugu beti mugimendu adierazgarriak txertatzen hitz egiten dugun bitartean. Hori erabakitzeko probabilitate bat finkatu beharko da eskuz (ikus aurrerago [3.4](#) atala).

3.2.2 Hautatutako terminoak eta keinuak

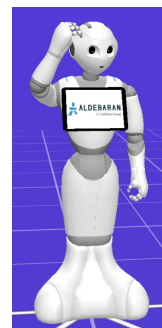
Hitz gako bakoitzari mugimendu bat edo bi esleitu zaizkio, eta antzeko esanahia duten hitzei mugimendu berdinak esleitu zaizkie, adibidez, "me", "I", eta "my" hitzei mugimendu berdinak esleitu zaizkie, hiruek hizlariari egiten baitiete erreferentzia. Hautatutako hitz gakoak hauek izan dira: Hello, Hi, I, me, my, Energy, Listen, Head, Not understand, Plenty, Spheres eta Come. 3.3. irudian termino bakoitzarentzat Chrograph-en jasotako postura desberdinak azaltzen dira.



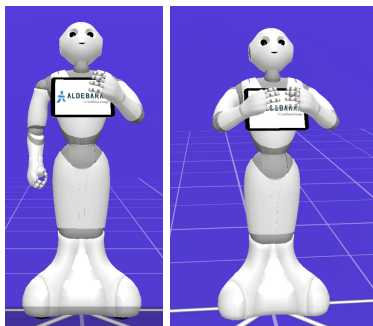
(a) Hello, Hi



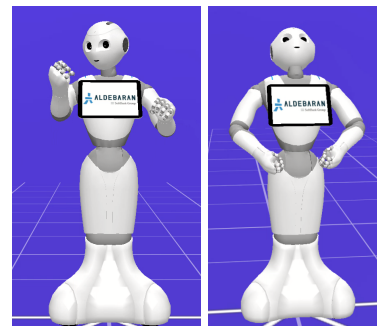
(b) Listen



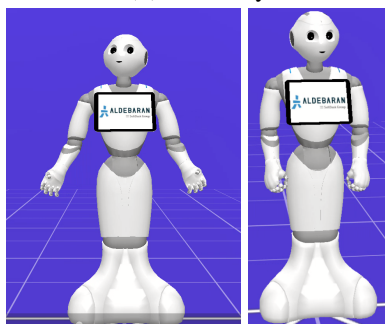
(c) Head



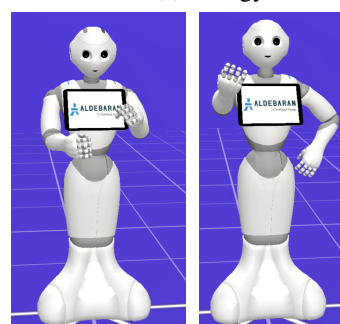
(d) I, me, my



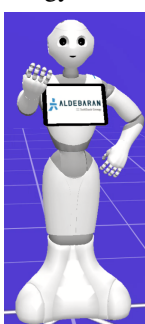
(e) Energy



(f) Not understand



(g) Spheres



(h) Come

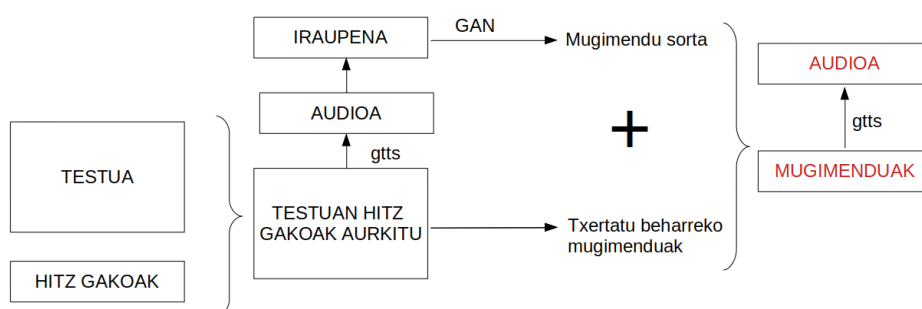
3.3 Irudia: Hitz gakoei lotutako mugimenduak

3.3 Txertatze prozesua: hurbilpen ezberdinak

Mugimenduen txertaketa egiteko bi hurbilpen erabili dira. Lehenean, mugimendua non txertatu aurkitzeko testuko hitzak erabiltzen dira, eta bigarrean, aldiz, testua audio bilakatu eta gero bilatzen da txertatze posizioa. Bakoitzak eskema propioa du, txertatzea ez baita momentu eta modu berdinean egiten.

3.3.1 Lehen hurbilpena

Gogora dezagun lehen urratsa robotak bota behar duen testuan hitz gakoak kokatzea dela. Lehen hurbilpen honetan, hitz gako horien bilaketarako eta kokapenerako 3.4 irudian erakusten den estrategia berezia diseinatu da.



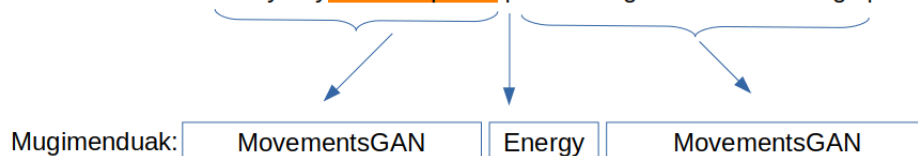
3.4 Irudia: Lehen hurbilpenaren eskema

Testua osotasunean edo esaldika tratatzea izan da hartu beharreko lehen erabakia. Testuaren luzeraren arabera, osorik prozesatuz gero sinkronizazio arazoak sor daitezke, bereziki hitz gakoaren mugimenduak luzeegiak baldin badira. Testu osoan zehar termino ugari baldin badago, sinkronizazio erroreak metatu egingo dira. Bestetik, testua esalditan banatu eta prozesamendua esaldika eginez gero, gerta daiteke mugimenduek naturaltasuna galtzea.

Hitz zerrenda (mugimenduei lotutakoa) eta robotak esan behar duen testua emanik, testua esaldika banatu eta prozesatu da. Intereseko termino bat topatuz gero termino horren aurretik zeuden 15 karaktereak gorde eta hitz gakoa bera 'l' karaktereaz ordeztu da banatzaile gisa. 15 karaktere horiek hitz gakoa identifikatzeko erabiliko dira, hitz gako bakoitzaren aurretik dauden 15 karaktereak ezberdinak direla suposatuz. Izan ere, terminoak

Hasierako testua: every day I wake up with **energy** and willing to learn new things.

Hitza aurkitu eta: every day **I wake up with** | and willing to learn new things |



Audioa: every day I wake up with energy and willing to learn new things.

3.5 Irudia: Lehen hurbilpena

banatzaileaz ordezkatu direnez, 15 karaktereei esker haien posizioa gordetzen da, ondoren txertaketa egin ahal izateko. Banatzaile hori bera esaldiaren amaieran ere gehitu da, amaiera bera adierazteko eta eten puntuak sortzeko eta sinkronizazio arazoak leuntzeko. Testua prozesatu ondoren, bi esaldi edo segmentu mota dauzkagu: hitzari lotutako mugimendua txertatzea eskatzen dutenak eta bestelakoak. Bi moten artean desberdintzeko, segmentuaren azken 15 karaktereak aztertu behar dira, eta aurretik gordetakoan artean baldin badaude, hitz gakoa eta dagokion mugimendua hor bertan txertatu beharko dira. Segmentu bakoitzarentzat GAN sareari eskaera desberdina egingo zaio beraz.

Laburbilduz, lehen hurbilpenean hitz gako bat aurkitzean 'l' banatzaileaz ordezkatu da eta aurretik dauden hamabost karaktereak gordetzen dira. Ondoren, GANak sortutako edo txertatu behar diren mugimenduak toki egokian jartzen dira eta hitzak eta mugimenduak elkarrekin exekutatu dira. 3.5. irudian eskema simple batekin ikusten da testuko lagin bat nola prozesatu den, hitz gakoa "energy" delarik. Adibide honetan GAN sareari bi aldiz deitzen zaio orduan.

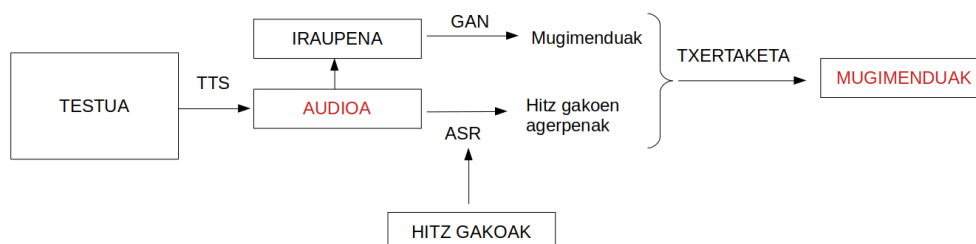
Exekuzio garaian arazoak aurkitu dira mugimendu luzeak eskatzen dituzten terminoekin. Termino horiek ahoskatzeko behar den denbora, mugimenduak exekutatzeko baina luzeagoa baitzen. Hori dela eta, *Choregraphe* erabili da eta mugimendu horiek egokitu dira iraupena egokitzeko.

Bestalde, mugimenduak txertatzeko momentuari buruzko informaziorik ez daukagu. Hau da, hitz gakoak eta haien mugimenduak non txertatu jakiteko 15 karaktere horiek esperimentalki aukeratu dira. Hortaz, testuan zehar bi esaldi antzekoak badira eta bi hitz gakoren aurretik dauden 15 karaktereak berdinak baldin badira, gerta liteke hitz egokia ez txertatzea, karaktere horiei esker baitakigu zein hitz gako txertatu behar den. Denbora

etiketak izanez gero, txertatze prozesua zehatzagoa izango litzateke eta, hortaz, portaera, sinkronizazioari dagokionez behintzat, egokiagoa.

3.3.2 Bigarren hurbilpena

Bigarren hurbilpenean, *IBM Watsoneko Text To Speech (TTS)* eta *Automatic Speech Recognizer (ASR)* tresnak erabili dira robotak esango duen audioa sortzeko eta audio horretan hitz gakoak bilatzeko, hurrenez hurren. Watson-en TTS-ak sortutako audiotik, modu zehatzean esaldiaren luzera zein den jakin daiteke, terminaleko media info aginduari esker. ASRari esker, hitz gakoak zerrenda pasata hitz bakoitzaren agerpenen lehen eta azken indizeak eskuratzen dira. Indize horiek segundoak dira eta audioaren hasierarekiko hitzak noiz hasi eta bukatzen diren adierazten dute. Horrela, mugimendua non txertatu jakiteko behar dugun informazio guztia daukagu. 3.6 irudian laburbiltzen da prozesu hau guztia.

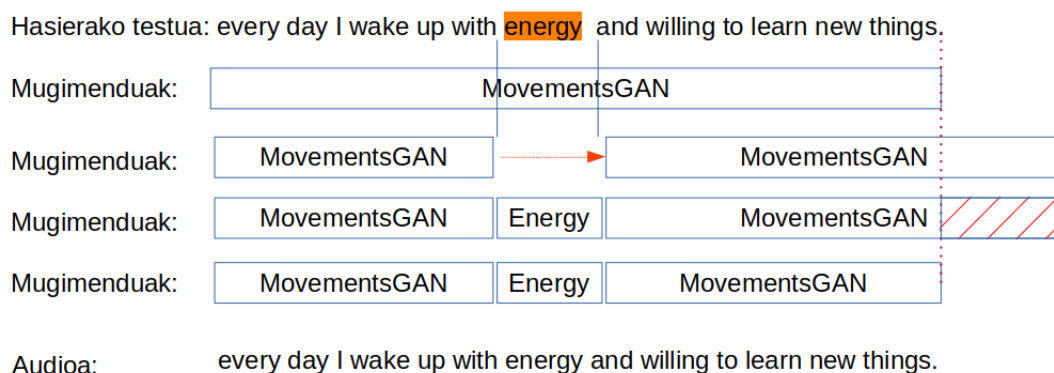


3.6 Irudia: Bigarren hurbilpenaren eskema

Hitz bat baino gehiago aurkitzen diren kasuan, ausaz erabakiko da zein hautatuko den txertatzeko, betiere probabilitate handiagoa emanez gutxien agertzen diren hitzei.

GANak audioaren iraupen osorako mugimenduak sortuko ditu. Oraingoan, *Watsonek* emandako denbora unetan txertatu beharko dira mugimendu berriak, eta ondoren datozenak desplazatu. Honek, noski, mugimendu segida luzeegia sortuko du eta esaldi amaierako zenbait mugimendu ezabatzea eskatzen du, beti ere hitz gakoa ez bada azken posizioan dagoena. Hitz gakoa esaldi bukaeran dagoenean bi aukera daude, alde batetik, ez moztea, hau da, nahiz eta hitz egiten bukatzen duen, mugimendua bukatu arte itxoitea, edo mugimendua aurreratzea, lehenago egiten hastea. Horrela, diskurtsoak eta mugimenduek elkarrekin bukatuko dute eta txertatutako mugimendua osoki egingo da. Bigarren hori izan da aukeratutako estrategia, mugimendua lehenago egiten hastea, esaldi bukaeran mugimendu

gehigarriak egitea baino naturalagoa gertatzen baita. Lehen hurbilketan erabilitako lagin berdinarekin, bigarren hurbilketaren prozesua ikus daiteke 3.7. irudian.



3.7 Irudia: Bigarren hurbilpena

Hurbilpen honek baditu zenbait koska. Batetik, *Watson* tresnak denbora behar du eskaera prozesatzeko, eta honek sistema desegokia bihurtzen du zuzeneko elkarrizketak eskatzen dituen kontesturako. Gainera, dohaineko bertsioak erabilera mugak ditu, *ASR*arentzat 500 minutuko prozesamendua hilabetean eta *TTS*arentzat 10000 karaktere hilabetean. Bestetik, audioan zuzenean hitz gakoak bilatzeko erabilitako *ASR*-ak arazoak erakutsi ditu zenbait hitz topatzeko, intonazioagatik edo puntuazio ikurrengatik antzaz.

3.4 Mugimenduak txertatzeko probabilitatea

Lehen aipatu bezala, gizakiok ez diogu sistematikoki hitz bati mugimendu konkretu bat lotzen, eta ideia robotak gauza berdina egitea da. Horretarako, hasiera batean hitz bakoitzari mugimendu bakarra lotu zitzaion, orduan ausaz hautatu zitekeen agerpen bakoitzean ea mugimendua egin edo ez. Baina gizakiok ez diogu hitz bati mugimendu bakarra lotzen, adibidez, norbait agurtzeko ez dugu beti jestu berdina egingo. Hortaz, robotean hori inplementatu da; hitz bakoitzari mugimendu bat baino gehiago esleitu zaio, eta txertatzerakoan ausaz hautatu da zein egin.

Esaldi batean hitz gako bat baino gehiago agertuz gero, bakar bat aukeratuko da, ausaz, baina probabilitate baten arabera, hitz gakoek testu arruntetan duten agerpen maiztasunak desberdinak dira eta. Honek, hitz gako batzuen presentzia areagotzen du eta robotaren

portaera errepikakor izatera bultza dezake. Izan ere, hitz gako batzuk testuan besteak baino askoz gehiagotan agertzen dira, eta orduan dagokien mugimenduak gutxiagotan egitea komeni da, bestela mugimendu horiek gehiegi egiten dira. Gainera, mugimendu batzuk besteak baino esanguratsuagoak dira, eta horiei ere pisu handiagoa eman zaie. Horregatik, hitz gako bakoitzari hautatua izateko probabilitate bat esleitu zaio. Esaldian agertzen den hitz gako kopuruaren eta bakoitzaren pisuaren arabera kalkulatu da hitz bakoitzak hautatua izateko probabilitatea, 0 eta 1 artean normalizatuz. Bestalde, pisuak estatikoak dira, hitz bat esaldi batean egoteak ez du esan nahi hurrengo esaldian egoteko probabilitatea txikituko zaionik. Gainera, pisuak edozein balio numeriko izan daitezke, ondoren normalizatzen baitira.

3.5 Sinkronizazioa

Proiektuaren garapenean zehar izandako arazo nagusia ez da mugimendua non txertatu jakitea, baizik eta txertatu nahi den mugimenduaren luzera kudeatzea. Mugimendu eta hitzen arteko sinkronizazioa bermatzea oso garrantzitsua da, bestela mugimendu esanguratsuak egiteak ez du ezertarako balio, ez badira momentu egokian egiten.

Lehen aipatu bezala, testuaren analisia egiteko eta hitz gakoak bilatzeko bi estrategia daude: testua esaldika edo kolpe batez prozesatzea. Sinkronizazioaren aldetik ikusia, esaldika egiten bada eta gakoari dagokion mugimenduaren luzera hitzarena berarena baina askoz luzeagoa baldin bada, esaldi bakoitzaren bukaeran soberan dauden jestuak egingo dira. Bestetik, testuaren mugimenduak jarraian exekutatzeko bada, errorea metatzen da. Horren ondorioz, robotak testua esaten bukatu duenean, metatutako erroreari dagokien mugimenduak egingo ditu, ezer esan gabe. Hori dela eta, testua esaldika egikaritzea izan da aukeratutako bidea, sinkronizazioa egokitzeko asmoz.

Bestetik, txertatu nahi den mugimendua egiteko behar den denbora hitza esateko behar dena baina askoz luzeagoa bada, nahiz eta mugimendua behar den momentuan txertatu, sinkronizazio arazo bat gerta daiteke, hitz gakoaren ahoskatzea mugimenduaren exekuzioa baina askoz lehenago bukatuko baita. Arazo hau bakarrik zentzu honetan gertatzen da, hots, mugimenduak nahiko luzeak izanik, aurkako egoera ez da gertatu, ez da hitz bat esateko behar den denbora baina mugimendu motzagorik gertatu, robotak denbora behar baitu mugimenduak gauzatzeko eta, horretarako erabiltzen diren serbomotorrak motelak baitira.

Gainera, hitza oso motza baldin bada mugimenduari konparatuz, hitz hori esan eta gero

mugimenduaren exekuzioak denboran iraungo du, eta hori ez da guk nahi duguna. Helburua hitza eta mugimendua elkarrekin egitea da, eta ez hitza esan eta gero mugimenduak segitzea.

Hortaz, bi kasuetan sinkronizazio arazoak ekiditeko garrantzitsua da hitzen eta mugimenduen luzeraren artean tarte handirik ez egotea. Horrela, hitza ahoskatu eta mugimendua exekutatu batera egiten da eta mugimendu gehiegi soberan ez izatea bermatzen da. Hala ere esaldi edo testu bukaeran mugimendu sobera badago, horiek kendu behar dira, testua eta mugimenduak elkarrekin bukatzeko. Hitz gakoa esaldi bukaeran baldin badago, ez da ezabatu behar noski eta txertaketa lehenago hasi behar da, mugimendua justu-justu esaldi bukaeran egiteko eta mugimenduaren bukaera hitzaren bukaera elkarrekin gertatzeko.

Hori guztia kontutan hartuta, sinkronizazio egokia ziurtatzeko mugimendu motzei lotutako hitzak hautatzea erabaki dugu. Horrela, mugimendua egiteko eta hitza esateko denboraren artean ez dago tarte oso handirik. Gainera esaldiak nahiko luzeak dira (10-15 hitz baino gehiagokoak ingelesez), eta horrela txertatzen den mugimendu bat ez baldin bada oso motza, mugimenduetan atzerapen bat sortzen bada errekueratu ahal izango da.

4. KAPITULUA

Esperimentazioa eta emaitzak

Implementazioa egin eta gero, robotean proba batzuk egin dira portaera globala ebaluatze-ko asmoarekin. Roboteran eramandako estrategia 2.a izan da, hau da, *Watson*ek eskaintzen digun denbora etiketak erabiltzen dituen estrategiarekin. Robotari emandako testua hau izan da, eta gorriz hitz gakoak non dauden ikus daiteke:

`Hello` everybody, it has been a long time, but it is a pleasure to see you again. `I` am Pepper, a humanoid robot designed to interact socially with humans. For now `my` capabilities are limited, but every day `I` wake up with `energy` and willing to learn new things. `I` am able to `listen` to your voice through the microphones built into `my head`, try to ask `me` something. We can talk about different topics, but bear in mind that `I` may `not understand` your question. Besides that, `I` have `plenty` of sensors that help `me` to interact with the environment. `I` can recognize objects of different shapes, for example `spheres` so `I` will notice if you play football around `me`. If you want to know `me` better, `come` to visit `me` to the robotics laboratory at the Faculty of Informatics in Donostia.

Testuak zortzi esaldi ditu eta 3.2.2. atalean definitutako hitz gakoak hogeit hamar aldiz agertzen dira. Hitzak testuan bilatzeko garaian, maiuskulek eta minuskulek ez dute garrantzirik, berdin-berdin aurkituak izango dira, *IBM Watson*eko *ASR*ak audioarekiko aurkitzen baititu hitz gakoak testuan.

Probetan erabilitako hitz gako bakoitzak duen pisua 4.1. taulan agertzen da. 3.4. atalean azaltzen den bezala, hitz gako bakoitzari lotutako pisu hauek esperimentalki finkatu dira, eta goiko adibideko testua irakurtzean nabaritzen da pisuen ezberdintasunaren arrazoia. Izan ere, "I", "me" eta "my" hitzak besteak baino askoz gehiago agertzen direla ikus daiteke. Gainera, antzeko esanahia dutenez (hiruek hizlariari erreferentzia egiten diete), mugimendu berdinak esleitu zaizkie. Horregatik beharrezkoa da hitz horien hautatuak izateko probabilitatea txikitzea, bestela, kasu gehienetan mugimendu berdinak txertatuko lirатеke, eta testuan behin edo bitan agertzen diren hitz gako mugimenduak ez lirатеke ia inoiz egingo.

Hitza	Pisua
Hello	80
I	1
Energy	70
Listen	60
Head	70
Me	1
Not understand	80
Plenty	70
Spheres	80
Come	80
My	1
Hi	70

4.1 Taula: Hitz gakoak eta haien pisuak

Egindako esperimentuekin hiru bideo grabatu dira, proiektuan zehar garatutako lorpenak ikusi ahal izateko:

1. *Txertatu_gabe.mp4* bideoak ¹, proiektua hasi baino lehen Pepper-ek egiten zituen mugimenduak agertzen dira, bakarrik *MovementsGAN* modulua erabiliz. Erreferentziarako bideoa da hau, proiektuan egindako lanak eragindako portaera aldaketa nabarmendu ahal izateko.

¹https://drive.google.com/file/d/14F_UlE-E1zGE9q87dXrA48qvy2uIYlMU/view?usp=sharing

Esaldia	Bigarren proba	Hirugarren proba
1. esaldia	Hello	Txertaketarik ez
2. esaldia	I	Txertaketarik ez
3. esaldia	energy	Txertaketarik ez
4. esaldia	head	head
5. esaldia	not understand	not understand
6. esaldia	plenty	plenty
7. esaldia	I	me
8. esaldia	come	come

4.2 Taula: Bi eta hirugarren probetan txertatutako mugimenduak

2. *Beti_txertatuz.mp4* bideoan ², hitz gakoien probabilitateak kontuan hartu gabe robotak erakusten duen portaera ikus daiteke. Termino bat topatuz gero, beti egiten du mugimendua. Txertatutako mugimenduak esanguratsuak dira eta esandako esaldiei zentzu gehiago ematen die, bakoitzean hitz bat mugimenduekin ilustratzen baita. Hala ere, kontuz ibili behar da txertatutako mugimendu kopuruarekin, gizakiok ez baitugu bortxaz esaldiak hitz baten ilustrazioa egiten gure mugimenduekin.
3. *Batzutan_txertatuz.mp4* bideoan ³, bertsio egokiena erakusten da, mugimendua txertatu edo ez ausaz aukeratzen da, gertatzeko %50eko probabilitatearekin. Hor bai gizakion portaerara hurbiltzen garela, bakarrik esaldi batzuetan txertaketa egiten baita. %50eko probabilitate hori aldatzea badago, txertatuak izatea nahi dugun mugimendu kopuruaren arabera, gehiago nahi baditugu zenbakia handitu beharko da eta noizean behin bakarrik txertatzea nahi badugu txikitu beharko da.

4.2. taulan ikus daitezke bi eta hirugarren probetan esaldi bakoitzean zein hitz gako hautatua izan den.

Bideo batetik bestera, ikusten da naturaltasuna irabazten dela, eta gizakiok dugun portaerara hurbiltzea lortu dela.

²<https://drive.google.com/file/d/13cExAoDD3HGGVUdWg5eqyfKGBoknLZbr/view?usp=sharing>

³<https://drive.google.com/file/d/11ATczLHA7W4vcdYBYkf2v39npzCbttUJ/view?usp=sharing>

3. ATALA

Ondorioak eta etorkizunerako lana

5. KAPITULUA

Ondorioak eta etorkizuneko lanak

Azken kapitulu honetan lanetik ateratako ondorioak eta etorkizuneko lan ideia batzuk aurkezten dira.

5.1 Ondorioak

Proiektu honen helburua robotaren gestikulazio arruntean semantikarekin lotutako gorputz mugimenduak txertatzea da. Horretarako, soluzio desberdinak planteatu eta aurrera eraman dira.

Emaitzetako bideoetan ikusten da mugimendu adierazgarriak txertatu eta naturaltasuna irabazi duela robotak, gizakiok egiten ditugun mugimenduetaz hurbiltzen baikara.

Robotaren diskurtsoan behar ziren mugimenduak txertatzea lortu da, behar zen momentuan eta esaldi bukaeran mugimendu gehigarriarik egon gabe. Gainera, ausazkotasun pixka bat sartu da eta askotan exekutatu gero, ez du beti hitz berdinetan mugimendua egiten eta hitz berdinean egiten badu, ez du beti mugimendu berdina egiten. Horri esker, hasierako emaitzetan genuena baino naturaltasun handiagoa ukaitea lortu dugu.

Hortaz, ikusten da sare neuronalik erabili gabe diskurtsoa laguntzen duten mugimendu naturalak egiteko aukera dagoela.

Hala ere, robotaren portaerak limitazio batzuk erakutsi ditu. Izan ere, hitz kopuru mugatu batentzat bakarrik erabili daiteke, eta testu zehatz bati egokitzen zaio, hortaz, testuz alda-

tu nahi badugu, beste hitz eta mugimendu batzuk hautatu beharko dira, eta bestela, hitz orokor asko jarri beharko dira, edozein testurako balio dutenak.

Gainera, momentuz hitz gako bakoitzari mugimendu bat edo bi esleitu zaizkio, eta oraindik naturaltasun gehiago lortzeko, mugimendu gehiago aurkitu beharko lirateke, horrela hitz orokor asko izanda ere, ez lirateke beti mugimendu berdinak txertatuko.

5.2 Etorkizuneko lanak

Lan honi jarraipena emateko, esaldi batean hitz gako asko dauden kasuan, hitz bat baino gehiagori lotutako mugimenduak egitea ideia bat izan daiteke; edo hitzak hautatuak izateko probabilitateak dinamikoki egokitzea, hau da, eskuz sartu behar ez izatea, baizik eta jadanik txertatuak izan diren mugimenduei lotutako hitzen probabilitatea jaistea, edo testu osoan zehar hitz bakoitzaren agerpen kopuruaren arabera probabilitateak kalkulatzeko.

Horrez gain, txertaketa "eskuz"egin ordez ikasketa sakonari esker esanahi semantikoa duten mugimenduak egiten lortzea ere interesgarria izan daiteke. Gainera, garatutako sistemen emaitzen ebaluazioa egiteko metodoak sortu edo aplikatzea ere onuragarria izango litzateke.

Bukatzeko, robota norekin hizketan ari den ezagutzeko gai bada, hau, da, aurpegiak ezagutzeko gai baldin bada, bere solaskidearen arabera mugimendu ezberdinak egitea interesgarria izan daiteke, betiere gizakion portaeraz hurbiltzeko asmoz, gizakiok ez baititugu gure solaskide guztiekin mugimendu adierazgarri berdinak egiten.

Eranskinak

Esteka interesgarriak

- [Pepper robotari buruzko dokumentazioa](#)
- [Choregrapheri buruzko dokumentazioa](#)
- [NAOqi SDK](#)
- [Softbank Robotics](#)
- [Hanson Robotics](#)
- [NAOqi Framework](#)
- [IBM Watsoneko Text To Speech](#)
- [IBM Watsoneko Speech To Text](#)
- [ALMotion APIaren dokumentazioa](#)
- [gTTS dokumentazioa](#)
- [Pydub liburutegia](#)
- [Robot ezberdin pila bat](#)
- [Martitzeko Perseverance roverrari buruzko informazioa](#)
- [Nuka robotari buruzko informazio gehigarria](#)

Bibliografia

- [Alexanderson et al., 2020] Alexanderson, S., Székely, É., Henter, G., Kucherenko, T., and Beskow, J. (2020). Generating coherent spontaneous speech and gesture from text. *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*.
- [Bressemer and Ladewig, 2011] Bressemer, J. and Ladewig, S. (2011). Rethinking gesture phases: Articulatory features of gestural movement? *Semiotica*, 184:53–91.
- [Cabibihan et al., 2013] Cabibihan, J.-J., Javed, H., Jr, M., and Aljunied, S. (2013). Why robots? a survey on the roles and benefits of social robots in the therapy of children with autism. *International Journal of Social Robotics*, 5.
- [Calo et al., 2011] Calo, C. J., Hunt-Bull, N., Lewis, L., and Metzler, T. (2011). Ethical implications of using the paro robot with a focus on dementia patient care. In *Proceedings of the 12th AAAI Conference on Human-Robot Interaction in Elder Care, AAAIWS'11-12*, page 20–24. AAAI Press.
- [Daily et al., 2017] Daily, S. B., James, M. T., Cherry, D., J. Porter, J., Darnell, S. S., Isaac, J., and Roy, T. (2017). Chapter 9 - affective computing: Historical foundations, current applications, and future trends. In Jeon, M., editor, *Emotions and Affect in Human Factors and Human-Computer Interaction*, pages 213–231. Academic Press, San Diego.
- [Diamond et al., 2012] Diamond, A., Knight, R., Devereux, D., and Holland, O. (2012). Anthropomorphic robots: Concept, construction and modelling. *International Journal of Advanced Robotic Systems*, 9(5):209.
- [Farley et al., 2020] Farley, K. A., Williford, K. H., Stack, K. M., Bhartia, R., Chen, A., de la Torre, M., Hand, K., Goreva, Y., Herd, C. D. K., Hueso, R., Liu, Y., Maki, J.Ñ., Martinez, G., Moeller, R. C., Nelessen, A., Newman, C. E., Nunes, D., Ponce, A.,

- Spanovich, N., Willis, P. A., Beegle, L. W., Bell, J. F., Brown, A. J., Hamran, S.-E., Hurowitz, J. A., Maurice, S., Paige, D. A., Rodriguez-Manfredi, J. A., Schulte, M., and Wiens, R. C. (2020). Mars 2020 mission overview. *Space Science Reviews*, 216(8):142.
- [Goodrich and Schultz, 2007] Goodrich, M. A. and Schultz, A. C. (2007). Human-robot interaction: A survey. *Found. Trends Hum.-Comput. Interact.*, 1(3):203–275.
- [Kucherenko et al., 2020] Kucherenko, T., Jonell, P., van Waveren, S., Henter, G. E., Alexandersson, S., Leite, I., and Kjellström, H. (2020). Gesticulator: A framework for semantically-aware speech-driven gesture generation. *Proceedings of the 2020 International Conference on Multimodal Interaction*.
- [Le et al., 2011] Le, Q. A., Hanoune, S., and Pelachaud, C. (2011). Design and implementation of an expressive gesture model for a humanoid robot. In *2011 11th IEEE-RAS International Conference on Humanoid Robots*, pages 134–140.
- [Mois and Beer, 2020] Mois, G. and Beer, J. M. (2020). Chapter 3 - robotics to support aging in place. In Pak, R., de Visser, E. J., and Rovira, E., editors, *Living with Robots*, pages 49–74. Academic Press.
- [Pandey and Gelin, 2018] Pandey, A. K. and Gelin, R. (2018). A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics Automation Magazine*, 25(3):40–48.
- [Retto, 2017] Retto, J. (2017). Sophia, first citizen robot of the world.
- [Scoglio et al., 2019] Scoglio, A., Reilly, E., Gorman, J., and Drebing, C. (2019). Use of social robots in mental health and well-being research: Systematic review. *Journal of Medical Internet Research*, 21:e13322.
- [Sheridan and Verplank, 1978] Sheridan, T. and Verplank, W. (1978). *Human and Computer Control of Undersea Teleoperators*. Man-Machine Systems Laboratory, Massachusetts Institute of Technology.
- [Stiefelhagen et al., 2004] Stiefelhagen, R., Fugen, C., Gieselmann, R., Holzappel, H., Nickel, K., and Waibel, A. (2004). Natural human-robot interaction using speech, head pose and gestures. pages 2422 – 2427 vol.3.
- [Thomaz et al., 2016] Thomaz, A., Hoffman, G., and Cakmak, M. (2016). Computational human-robot interaction. *Foundations and Trends in Robotics*, 4:104–223.

- [Wagner et al., 2014] Wagner, P., Malisz, Z., and Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57:209–232.
- [Yoon et al., 2020] Yoon, Y., Cha, B., Lee, J.-H., Jang, M., Lee, J., Kim, J., and Lee, G. (2020). Speech gesture generation from the trimodal context of text, audio, and speaker identity. *ACM Transactions on Graphics*, 39(6):1–16.
- [Yoon et al., 2019] Yoon, Y., Ko, W.-R., Jang, M., Lee, J., Kim, J., and Lee, G. (2019). Robots learn social skills: End-to-end learning of co-speech gesture generation for humanoid robots. In *Proc. of The International Conference in Robotics and Automation (ICRA)*.
- [YU and Tapus, 2020] YU, C. and Tapus, A. (2020). SRG3: Speech-driven Robot Gesture Generation with GAN. In *16th International Conference on Control, Automation, Robotics and Vision*, Shenzhen (virtual), China.
- [Zabala et al., 2019] Zabala, U., Rodriguez, I., Martínez-Otzeta, J. M., and Lazkano, E. (2019). Learning to gesticulate by observation using a deep generative approach. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- [Šabanović et al., 2013] Šabanović, S., Bennett, C. C., Chang, W.-L., and Huber, L. (2013). Paro robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. In *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, pages 1–6.